





Genome version 5.1.3  
Copyright (c) 1991-2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on January 31, 2003, 07:44:12 ; Search time 11 seconds  
(without alignments)  
1040 679 Million cell updates/sec

Title: US\_09\_855\_294b\_1

Perfect score: 1438

Sequence: 1 MTSFPIALQLAIISSIKSN

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying the search criteria: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

# SUMMARIES

Result No.	Score	Query length	DB ID	Description
1	773	53-8	1 YAU8_GCHPO	Q10166 schizosacch
2	694	48-3	1 YLAE_YEAST	P49964 saccharomyc
3	406	28-3	1 YLAE_YEAST	P47016 saccharomyc
4	392	27-3	1 YLAE_YEAST	P55175 synchocyst
5	324	22-5	1 Y480_MYCTU	Q11146 mycobacteri
6	283	19-7	1 Y480_MYCTU	P55177 staphylococ
7	263	19-3	1 YBEM_ECOLI	P19874 escherichia
8	257	17-9	1 YBEM_ECOLI	P55178 staphylococ
9	256	17-8	1 YBEM_ECOLI	P58064 escherichia
10	254	17-7	1 NPL2_ARATH	P12962 arabidopsis
11	247	17-2	1 YHGX_BACSU	P54608 bacillus su
12	244	17-0	1 NPL2_ARATH	P12961 arabidopsis
13	240	16-7	1 NPL2_ARATH	P46010 arabidopsis
14	239	16-6	1 YPOQ_PSEEL	P55176 pseudomonas
15	206	14-3	1 NPL2_ARATH	P42965 nicotiana t
16	203	14-0	1 NPL2_ARATH	Q03217 rhodospirillum
17	202	14-1	1 NPL2_ARATH	P46011 arabidopsis
18	175	12-2	1 ALAM_PSEAE	P11436 pseudomonas
19	173	12-0	1 YAFV_ECOLI	Q47679 escherichia
20	162	11-3	1 ALAM_PSEAE	Q41360 rhodospirillum
21	155	10-9	1 NPL2_ARATH	P62968 rhodospirillum
22	153	10-7	1 YLAE_YEAST	Q05498 homo sapien
23	152	10-6	1 YLAE_YEAST	Q03248 rattus norv
24	152	10-6	1 YLAE_YEAST	Q03248 rattus norv
25	151	10-5	1 YLAE_YEAST	Q03248 rattus norv
26	148	10-3	1 YLAE_YEAST	Q03248 rattus norv
27	149	10-3	1 YLAE_YEAST	Q03248 rattus norv
28	142	9-9	1 YLAE_YEAST	Q03248 rattus norv
29	140	9-6	1 YLAE_YEAST	Q03248 rattus norv
30	145	9-4	1 YLAE_YEAST	Q03248 rattus norv
31	134	9-3	1 YLAE_YEAST	Q03248 rattus norv
32	133	9-3	1 YLAE_YEAST	Q03248 rattus norv
33	129	9-0	1 YLAE_YEAST	Q03248 rattus norv

34	128.5	8.2	512	1 YLAE_YEAST	Q03248 rattus norv
35	128.5	8.2	567	1 YLAE_YEAST	Q03248 rattus norv
36	128.5	8.2	568	1 YLAE_YEAST	Q03248 rattus norv
37	123	8.6	552	1 YLAE_YEAST	Q03248 rattus norv
38	123.5	8.5	523	1 YLAE_YEAST	Q03248 rattus norv
39	119.5	8.2	567	1 YLAE_YEAST	Q03248 rattus norv
40	114	7.9	519	1 YLAE_YEAST	Q03248 rattus norv
41	105	7.3	199	1 YLAE_YEAST	Q03248 rattus norv
42	97.5	6.8	322	1 YLAE_YEAST	Q03248 rattus norv
43	97.5	6.8	558	1 YLAE_YEAST	Q03248 rattus norv
44	97.5	6.8	679	1 YLAE_YEAST	Q03248 rattus norv
45	97	6.7	555	1 YLAE_YEAST	Q03248 rattus norv

## ALIGNMENTS

RESULT 1  
YAU8\_GCHPO STANDARD; PRT: 322 AA.  
AC 210:56,  
DT 01-OCT-1996 (Ref. 34, Created)  
DT 01-OCT-1996 (Ref. 34, Last sequence update)  
DT 15 JUN 2002 (Ref. 41, Last annotation update)  
DE Hypothetical protein C26A3.11 in chromosome I.  
GN SPAC26A3.11.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
CC Schizosaccharomycetes; Schizosaccharomycetes;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
PP SEQUENCE FROM N.A.  
PC STPAIN-972;  
RX MEDLINE=2184801; PubMed=11859360;  
RA Wood V., Gilliam P., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
Brooks K., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
Holtz S., Holsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
McNey E., Moule S., Murgall K., Murphy L., Niblett D., Odell C.,  
Olliver P., O'Neil J., Pearson D., Seal M.A., Sefton D., Smith E.,  
Skelton J., Simmonds M., Squares P., Squares S., Stevens K.,  
Taylor V., Taylor P.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
Woodward J., Wolke A., Aert P., Pothier J., Grynolprez B.,  
Walters I., Vanstreels E., Pieger M., Schaefer M., Mueller-Auer S.,  
Gabel J., Fuchs V., Fritz C., Holzer E., Moesl U., Hilbert H.,  
Borzym K., Langer I., Beck A., Lebrach H., Reinhardt P., Pohl T.M.,  
Eger P., Giermann W., Wiedler H., Warburton F., Dunnally B.,  
Geffeau A., Cadieu E., Dreano S., Gloux S., Gloux V., Jolani S.,  
Galibert F., Aves S.J., Xiang Z., Hunt G., Moore K., Hurst S.M.,  
Lalzar M., Becker M., Tallard M., Tallard V.A., Guzik A., Thibault G.,  
Paga R.P., Choudhury R., Jimeur J., Sanchez M., A. P. P. P., Benard J.,  
Dominguez A., Revuelta J.L., Morero S., Armstrong J., Forburg S.L.,  
Cerrutti L., Lowe R., McCombie M.R., Paulsen O., Patashnik J.,  
RA Shpakovski G.V., Ussery D., Barrett P.G., Nurse P.,  
FT The genome sequence of Schizosaccharomyces pombe.  
PL Nature 415 971-980(2002)  
CC SIMILARITY: BLASTS TO THE UPF0012 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. See <http://www.isit.sib.ch/infocenter> or send an email to [license@sib.ch](mailto:license@sib.ch).  
EMBL: C09430; CAA0394.1;  
InterPro: IPR000000; N1567/Thy140

Query Match 48.3%; Score 694.5; D2 L1 Length 291,  
Best Local Similarity 49.8%; Freq No 1 qe-50;  
Matches 137; Conservative 52; Mismatches 77; Indels 9; Gaps 4;

5 RLALIQIQLSS-IGSDVNTKACSFIRPAATG--GAKIVSGFEFELSYGAKITFEYAE-- 59  
:  
12 KVALIVQISGGSPDKWANLQRAATHIERAKRQPTTKLVLLIEGENSYSTQDFRYGEVI 71  
:  
60 --KIPGEGTOKLSVAKECSLYILGGSGFEED-AGKLNYITCAVEPDTTLIAFPYTHL 115  
:  
72 NPKRGSTSVQSLSKAUKFKILLVGCTHELLEKTATVTNLTINREGFTEHPHYVHL 131  
:  
116 ELDIDVEGRKITFESKRTISTGSFSTFDTFICERVELGITCYDMFFARLAOLYAKPGCOLLVY 175  
:  
132 FLVDLPNGISFHESETLSEGESKSTETIKAGFGWGCYEEMPFKKAMLCAPFTAFANMII 191  
:  
176 PCAENLTIGPAIWELLQSRRAVEDGVVVATAGAPARDKASYVAVGHSTVAFVFWFLVAY 235  
:  
192 PSAFNVTIGPLHWHLLASRAVDNRVTVMLGSPARKLGSSVHAHGSLVWFDPQIVAPA 251  
:  
236 STEEAIVYSDDILKYLAIIPQQIPVFPRKRSDLYA 270  
:  
252 QEGSEIIVAEIDPEVIESFPPQAVPIIKRRFEDVIS 286  
: :

RESULT 3

YQM6 YEAST  
IID YQM6 YEAST STANDARD; PRT; 307 AA.  
AC P47016;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 347.6 kDa protein in SPB10-GCD14 intergenic region.  
GN YJLI26W OR J07706.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota, Fungi, Ascomycota, Saccharomycotina, Saccharomycetes,  
OCC Saccharomycetales, Saccharomycetaceae, Saccharomycetes.  
OX NCBI\_TaxId=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY1679;  
RF MEDLINE=97103775; PubMed=8948101,  
RA Cziepluch C., Koide E., Fujol A., Jauniaux J.-C.;  
RT "Sequencing analysis of a 40.2 kb fragment of yeast chromosome X  
reveals 19 open reading frames including URA2 (5' end), TRP1, PRS2,  
SPB10, GCD14, KPE1, PHO86, KCA3, ASF1, OCT7, GCF3, two tRNA genes,  
three remnant delta elements and a Ty transposon";  
RL Yeast 12:1471-1474(1996).  
CC -!- SIMILARITY: BELONGS TO THE UPF0012 FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements>) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

EMBL; Z49401; CAA82421.1;  
SD; S0003662; YJLI26W.  
DR InterPro; IPR03010; NtIsle/CNhydase.  
DR InterPro; IPR001110; UPF0012.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR SMART; PS01227; UPF0012; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 307 AA; 34693 MW; 4878/C43B1GA828E CRC64;

Query Match 28.3%; Score 406.5; D2 L1 Length 307;  
Best Local Similarity 31.6%; Freq NO. 1.5e-26;  
Matches 95; Conservative 63; Mismatches 106; Indels 37; Gaps 5;

5 RLALIQISSLKENVNTRACSFIRPAATG--GAKIVSGFEFELSYGAKITFEYAE-- 54



GenCore version 5.1.3  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on January 31, 2003, 07:52:23, Search time 13 seconds  
(without alignments)  
428,406 Million cell updates/sec

Title: US 09 855 294B 1  
RefSeq source: 1439  
Sequence: 1 MTSFELALQLQLGKSRSDN .....QIPVFRQK9SDLYAVEMKRP 276

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

100000 seqs, 2000000 residues

Total number of hits satisfying chosen parameters: 122226

Minimum DB seq length: 3  
Maximum DB seq length: 20000000  
Post processing: Minimum Match 84  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications AA\*

- 1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB\_PEP.\*
- 2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB\_PEP.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB\_PEP.\*
- 4: /cgn2\_6/prodata/1/pubpaa/US05\_PUBCOMB\_PEP.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB\_PEP.\*
- 6: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB\_PEP.\*
- 7: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB\_PEP.\*
- 8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB\_PEP.\*
- 9: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB\_PEP.\*
- 10: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB\_PEP.\*
- 11: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB\_PEP.\*
- 12: /cgn2\_6/prodata/1/pubpaa/US10\_PUBCOMB\_PEP.\*
- 13: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB\_PEP.\*
- 14: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB\_PEP.\*

Pred No is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	Pos	Description
1	370	25.7	266	9	US-09-738-626-6568 Sequence 6569, Ap
2	332	13.1	153	3	US-09-738-626-6568 Sequence 188, App1
3	261	14.0	366	10	US-09-972-186A-4 Sequence 4, App1
4	198.5	12.8	390	10	US-09-972-186A-2 Sequence 2, App1
5	191	11.3	346	10	US-09-751-208A-2 Sequence 2, App1
6	161	11.2	224	10	US-09-925-300-1632 Sequence 1632, Ap
7	159	11.1	337	10	US-09-751-208A-4 Sequence 4, App1
8	153	10.7	501	10	US-09-863-339-1 Sequence 1, App1
9	113	7.9	370	10	US-09-863-339-2 Sequence 2, App1
10	97.5	6.9	738	9	US-09-712-363-238 Sequence 238, App1
11	89	6.2	545	9	US-09-712-363-238 Sequence 5134, Ap
12	95	5.9	503	9	US-09-738-626-6568 Sequence 5184, Ap
13	80	5.6	476	10	US-09-817-184-2 Sequence 2, App1
14	80	5.6	727	10	US-09-445-023A-12 Sequence 12, App1
15	80	5.6	832	10	US-09-323-448B-4 Sequence 4, App1
16	80	5.6	968	9	US-10-163-316-7 Sequence 7, App1
17	78.5	5.5	1940	9	US-10-815-242-13273 Sequence 4, App1
18	78	5.4	291	10	US-09-815-242-13273 Sequence 13273, A
19	78	5.4	449	10	US-09-815-242-13273 Sequence 13273, A

20	78	5.4	452	10	US-09-815-242-13273 Sequence 6569, Ap
21	76.5	5.3	639	10	US-09-782-906-3 Sequence 2, App1
22	76	5.3	635	10	US-09-815-242-13273 Sequence 13273, A
23	75.5	5.3	1394	12	US-10-871-231-2 Sequence 2, App1
24	75.5	5.3	1353	10	US-09-751-208B-4 Sequence 99, App1
25	75	5.2	291	10	US-09-815-242-13273 Sequence 74, App1
26	75	5.2	163	10	US-09-925-300-1632 Sequence 40, App1
27	75	5.2	390	10	US-09-925-300-1632 Sequence 14, App1
28	75	5.2	512	9	US-09-925-300-1632 Sequence 254, App
29	75	5.2	512	9	US-09-815-242-13273 Sequence 254, App
30	75	5.2	545	9	US-09-978-208A-254 Sequence 254, App
31	75	5.2	545	9	US-09-978-208A-254 Sequence 254, App
32	75	5.2	545	9	US-09-978-208A-254 Sequence 254, App
33	75	5.2	545	9	US-09-978-208A-254 Sequence 254, App
34	75	5.2	545	9	US-10-176-758-58 Sequence 58, App1
35	75	5.2	545	9	US-10-176-758-58 Sequence 58, App1
36	75	5.2	545	9	US-10-176-758-58 Sequence 58, App1
37	75	5.2	545	9	US-10-176-758-58 Sequence 58, App1
38	75	5.2	545	9	US-10-176-758-58 Sequence 58, App1
39	74.5	5.2	276	10	US-09-815-242-13273 Sequence 1129, A
40	74.5	5.2	2862	10	US-09-742-659-5 Sequence 304, App
41	74	5.2	2862	9	US-09-404-669-304 Sequence 304, App
42	74.5	5.2	2862	10	US-09-742-659-5 Sequence 304, App
43	74.5	5.2	2862	10	US-09-742-659-5 Sequence 304, App
44	74	5.1	516	10	US-09-822-863-2 Sequence 2, App1
45	74	5.1	508	10	US-09-822-863-2 Sequence 2, App1

## ALIGNMENTS

### RESULT 1

US-09-738-626-6568  
Sequence 6569, Application US/09/738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAFARAWA, SATOSHI  
APPLICANT: MITSUUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIYO  
APPLICANT: OCHIAI, KEIJO  
APPLICANT: YOKOI, HAPUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENO, AHIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OKAZAKI, AYO  
TITLE OF INVENTION: NOVEL POLYMERIZABLES  
FILE REFERENCE: 243-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 99/193162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 99/240988  
FILING DATE: 2000-03-03  
NUMBER OF SEQ ID NOS: 7959  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6568  
LENGTH: 266  
TYPE: PPT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6568

Query Match: 25.7% Score: 770, DE: 0, Length: 266

Best Local Similarity: 31.5% Pos: 0, Score: 0

Matches: 86, Conservative: 53, Mismatches: 108, Indels: 26, Gaps: 6,

5 ELLIQUININAVTANQFERNATVIVLSENGEYVAFIEEAEIPIGE 64

5 ELLIQUININAVTANQFERNATVIVLSENGEYVAFIEEAEIPIGE 61

65 STLESEVAREEELLYLDSIP... EELIQUININAVTANQFERNATVIVLSENGEYVAFIEEAEIPIGE 116

Db 62 ESTAVKELAEJLVAVIAGMFTADTVQGEKTSVNNNTVLISG--AGLHOGYNKHTY 119  
QY 117 EIDVPGKIPGSGKTSIPGSGSTPNPYFVGIIGICYMPPFAELAGIYVAGKCOLLVYP 176  
Db 120 P--APYPSSTVPGDHEIIVVFEVDIDIFFGVATYDIRFPEQPKDLARNGAQLIIVP 174  
QY 177 GAFNLTGTGA--HWELLQSPAVDQVNVVATASPA-----RDDKASYVWAGHSTVVP 227  
Db 175 TSMQ--DGPQKQLVAVLPAPALLSTWIVAGCGAPLPBELRCEKRGFTG:GHSMTNP 232  
QY 228 WESVLAFATHEAIVYSDIP:KYLAIEPQCLPV 260  
Db 233 HGEVIASAGVEPEMILADIDVSLGLAKIREALPV 265

RESULT 2  
US-09-974-639-188  
; Sequence 188, Application US/09774639  
; Publication No. US20020137153A1  
; ORGANISM: Homo sapiens  
; FEATURE: INFORMATION  
; OTHER INFORMATION: Description of Proteins  
; TITLE OF INVENTION: An Human Secreted Proteins  
; FILE REFERENCE: P033F1  
; CURRENT APPLICATION NUMBER: US/09/774,639  
; PRIOR FILING DATE: 2001-07-09  
; PRICE APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/244,112  
; PRICE FILING DATE: EARLIER FILING DATE: 1999-02-04  
; NUMBER OF SEQ ID NOS: 371  
; SOFTWARE: PatentIn Ver 2.0  
; SEQ ID NO 188  
; LENGTH: 153  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-774-639-188

Query Match 23.1%; Score 332.5; DB 9; Length 153;  
Best Local Similarity 45.2%; Pred No 2 56-26;  
Matches 66; Conservative 25; Mismatches 48; Indels 7; Gaps 3;  
QY 128 ESYTSLGDSF STEFTYFVGIYGVGIGIYDMPFAELAGIYVAGKCOLLVYPKAEMLTGP 186  
Db 3 ESNSTMPGSLSPVSTFAGIYGVGIGIYDMPFAELAGIYVAGKCOLLVYPKAEMLTGP 62  
QY 197 HWELLQSKAVKQVIVVATASPA---ELDKASYVWAGHSTVVPWGEVLAVAGTTEALVY 243  
Db 63 HWELLQSKAVKQVIVVATASPA---ELDKASYVWAGHSTVVPWGEVLAVAGTTEALVY 243  
QY 244 SDILYFLAEIYGVGIGIYGVGIGIYDMPFAELAGIYVAGKCOLLVYPKAEMLTGP 269  
Db 120 ARIDINYLQPLRHLVFPQHPRPDLY 145

RESULT 3  
US-09-972-186A-4  
; Sequence 4, Application US/09972186A  
; Patent No. US20020137153A1  
; GENERAL INFORMATION:  
; APPLICANT: PAMER, SANDRA W.  
; APPLICANT: HUISMAN, GJALT  
; APPLICANT: MILLIS, JIM  
; APPLICANT: SHELDON, ROGER  
; APPLICANT: DEL CARDAYRE, STEPHEN  
; APPLICANT: COX, ANTHONY  
; APPLICANT: DAVIS, S. CHRISTOPHER  
; TITLE OF INVENTION: ENANTIOSELECTIVE PRODUCTION OF AMINO CARBOXYLIC ACIDS  
; FILE REFERENCE: 02-108110US  
; CURRENT FILING DATE: 2001-10-04  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence. Synthetic  
; OTHER INFORMATION: peptide sequence  
US-09-972-186A-4  
Query Match 14.0%; Score 201.5; DB 10; Length 366;  
Best Local Similarity 24.8%; Pred No. 1 56-12;  
Matches 80; Conservative 51; Mismatches 124; Indels 67; Gaps 15;

QY 3 SPRLALIQIQ-ISSIKSDNVTRACSFIREAATOCAGKIVSLPECF --- --NSPYG- 50  
Db 7 TPKVAAVQAPQVWFDAAKTVDKIVSIIAEAAANGCGLVAFPEVPIGYPVHIVWDSPLAG 66  
QY 51 AKYFEYAEK...IFESTQMLSEVAECETILIGTSIFEDAGKLYNTPAVPQFETTL 106  
Db 67 MAKFAVRYHENSUTMSPHVQKLLDAARHSIAVAVG-ISEKGGSLIMTQLIIDAGGQL 125  
QY 107 LAKYRKHLPLDIDVPGKILTHQESKILSPG--SESTFETFCVAGLICYD -MRFVELA 162  
Db 126 VAPEPEL-----PPTHVSPVTVGERSNS:SVYUM:FAKLAALNWEHF-TUTKYA 176  
QY 163 QIVAGKCOLLVYFG-----AF-----NLTSFAHWELLQSKAVKQVIVVATASPA 229  
Db 177 MYSMIEQVHVWASWPGMSLYQPEVPAFGVDAQLIA-----IRMYALEGQIFVVCV 228  
QY 210 -----RDKASYV--AWHSTVVPWGEVLAK- AGTEEAIVYSDIDKZLAEL 254  
Db 229 VTENHEFFCENEEOPLIGFGGFGFAPILGPNRGRDLAIPLAUEEGILYADIDLSATLA 288  
QY 255 RQOI-PVEPQKPSDLVAVEMKK 275  
Db 289 QQAADPVGHYSPPDVLNINFNQ 310

RESULT 4  
US-09-972-186A-2  
; Sequence 2, Application US/09972186A  
; Patent No. US20020137153A1  
; GENERAL INFORMATION:  
; APPLICANT: PAMER, SANDRA W.  
; APPLICANT: HUISMAN, GJALT  
; APPLICANT: MILLIS, JIM  
; APPLICANT: SHELDON, ROGER  
; APPLICANT: DEL CARDAYRE, STEPHEN  
; APPLICANT: TOBIN, MATTHEW  
; APPLICANT: COX, ANTHONY  
; APPLICANT: DAVIS, S. CHRISTOPHER  
; TITLE OF INVENTION: ENANTIOSELECTIVE PRODUCTION OF AMINO CARBOXYLIC ACIDS  
; FILE REFERENCE: 02-108110US  
; CURRENT APPLICATION NUMBER: US/09/972,186A  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: peptide sequence  
US-09-972-186A-2

Query Match 13.8%; Score 198.5; DB 10; Length 380;  
Best Local Similarity 24.5%; Pred No. 1 20-12;  
Matches 79; Conservative 52; Mismatches 124; Indels 67; Gaps 15;



Fri Jan 31 13:48:38 2003

Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraiishi, Yoshiko  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/09/294,871A  
APPLICATION NUMBER: US/09/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICANT: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/IP91/01696  
FILING DATE: 06-DEC-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/IP92/00739  
FILING DATE: 10-JUN-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 212602/1992  
FILING DATE: 10-AUG-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/IP93/01101  
FILING DATE: 05-AUG-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/PATENT NUMBER: 741274/121/AUPA  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399

TELEX: 904136  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-46  
Query Match  
Best Local Similarity 16.8%, Score 241, DB 2, Length 303;  
Matches 75; Conservative 56; Mismatches 106; Indels 40; Gaps 12;  
QY 21 VTRACSFIREAATGGAKIVSLFECFSPYAK-YFEYAE-----KIPFESTYISEV 72  
DB 24 VVKLLDMLTKAASKGANFIVFPLAULTIFFRWYFTLAELEDSFEYTEMCPVVRPLPEY 83  
QY 73 AKECST-YLIGGS--IFEEDAGKLYNCAVFGPGGTLAYKRIHLFTIVDGVYIQF-- 127  
DB 84 AAEGLGIGENLGYAEVLVVEGGVKKRPNTSLVDESCHIVGFYFTHL---PRHFFYFAY 198  
QY 128 -----ESKTLSPSD-SFSTFTTTCRVGSGICVDKFAELALIVACRQGMILVYFGAFN 180  
DB 139 RPFQHLKRYPEFCGLGFPVVDVDAAMMELNURKWPPEAWVWGLRGAELICGGUN 196  
QY 181 LTT-GPA-----HWELLORSRAVDNVVATASPARDDKASTVAKGSHSTVNDPWG 229  
DB 197 IPTHNSVEQRLHLSHLLSWAGSYQTSAMSAAGKAGMFFHVMILCHGCTIVATG 265  
QY 230 EVLAKAGT-EEAIVYSDIDLKLAEIPQLIVFVQKP 265  
DB 456 EIVALTITLDEVITAADVLDRCRELEKEHIFHFQHP 292

RESULT 3  
US-08-876-398A-46  
Sequence 46, Application US/08876398A  
Patent No. 6083752  
GENERAL INFORMATION:  
APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NANBA, Hirokazu  
APPLICANT: TAKANO, Masayuki  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECAPAMYLASE IMPROVED IN  
TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/IP93/01101  
FILING DATE: 05-AUG-1993  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC 1992



QY 868 TGTGTACAGAAATAGGACGATATGATTTCTTCAACATATCAACTCCCTATTAAATCTTT 927  
 Db 841 TGGGTTCACAGAAATAGGACGATATGATTTCTTCAACATATCAACTCCCTATTAAATCTTT 900

QY 928 AATGAAGAAAAAAA 942  
 Db 901 TATGAAAAAAA 915

RESULT 2  
 BL333425  
 LOCUS  
 DEFINITION  
 mRNA sequence.  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates, Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1. (bases 1 to 893)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Sequencing by: Incyte Genomics, Inc.  
 DNA Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM11342 low: m column. 12  
 High quality sequence stop: 874.

FEATURES  
 Location/Qualifiers  
 1..893  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:513251"  
 /clone\_lib="NIH MGC\_12"  
 /tissue\_type="cervical carcinoma cell line"  
 /lab\_host="DH10B"  
 /note="Organ: cervix; Vector: pCMV SP6F6; Site: 1. NotI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.4 kb. Library prepared by Life  
 Technologies."

BASE COUNT 245 a 202 c 212 g 233 t 1 others

ORIGIN  
 Query Match 61.0%; Score 828.4; DB 13; Length 893;  
 Best Local Similarity 97.6%; Pred. No. 2.7e-137;  
 Matches 872; Conservative 0; Mismatches 17; Indels 4; Gaps 3

QY 22 ATGACCTCTTTCGGCTTGGGCTTCCATCCAGCTTCAGATTCCTTCATCAATCAGATAAC 81  
 Db 1 ATGACCTCTTTCGGCTTGGGCTTCCATCCAGCTTCAGATTCCTTCATCAATCAGATAAC 60

QY 82 GTCACCTGGCTTGTAGCTTCATCCGGGAGGAGCAACGAGGAGCAAAATAGTTTCT 141  
 Db 61 GTCACCTGGCTTGTAGCTTCATCCGGGAGGAGCAACGAGGAGCAAAATAGTTTCT 119

QY 142 TTGCGGAAATGCTTTAAATCTCCATATGGAGGAAATATTTCTGAATGAGAGAAA 201  
 Db 120 TTGCGGAAATGCTTTAAATCTCCATATGGAGGAAATATTTCTGAATGAGAGAAA 179

QY 202 ATTCTGTGTGAATTCACACAGAGCTTCTGAGTAGCAAGGAATCCACATATATCTC 261  
 Db 180 ATTCTGTGTGAATTCACACAGAGCTTCTGAGTAGCAAGGAATCCACATATATCTC 239

QY 242 ATTGGAGCTCTATCCCTGGAAGAGATGCTGGGAAATATATACACCTGTGCTGTTT 321

Db 240 ATTGGAGGCTTATATGCTTAAAGAGATGTTGAGAAATATATATACAAAGCTGTGCTGTGTTT 244  
 QY 322 GAGCTGATATTAATTTATATAAATATATATATATATATATATATATATATATATATAT 326  
 Db 340 GAGCTGATATTAATTTATATAAATATATATATATATATATATATATATATATATATAT 344  
 QY 382 CTTGAGAAATTAATTTATATAAATATATATATATATATATATATATATATATATATATAT 441  
 Db 360 CTTGAGAAATTAATTTATATAAATATATATATATATATATATATATATATATATATATAT 419  
 QY 442 TTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 501  
 Db 420 TTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 479  
 QY 502 CTTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 561  
 Db 480 CTTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539  
 QY 562 CTTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 621  
 Db 540 CTTGATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 599  
 QY 622 GTATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 681  
 Db 600 GTATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659  
 QY 682 CACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 749  
 Db 660 CACAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719  
 QY 740 CAAATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 799  
 Db 720 CAAATATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 778  
 QY 800 TTTTATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 859  
 Db 778 TTTTATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 838  
 QY 860 GTTTTATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 912  
 Db 839 GTTTTATATCTTATATGACATATGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 891

RESULT 3  
 BG697375  
 LOCUS 602661148F1 NC1\_USAP\_Skn3 Homo sapiens cDNA clone IMAGE:4804451 5',  
 946 bp mRNA linear EST 07-MAY-2001  
 DEFINITION mRNA sequence  
 DESCRIPTION BG697375  
 VECTIN P527375.1 GI 139673.1  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 BUKARYOVA, Mariya, Charduta, Craniata; Vertebrata; Euteleostomi,  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 i. (bases 1 to 946)  
 NIH-MRC <http://imgc.mrc.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
 Tissue Procurement: James Cleaver, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA  
 Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: WGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLNL0701 row: o column: 12  
 High quality sequence stop: 839  
 Location/Qualifiers  
 1..846  
 /organism "Homo sapiens"

FEATURES  
 source

BASE COUNT 244 a 144 c 220 t  
 ORIGIN  
 Query Match 59.4%; Score 807.6; DB 12; Length 846;  
 Best Local Similarity 99.2%; Pred. No. 1.3e-133;  
 Matches 833; Conservative 0; Mismatches 4; Indels 3; Gaps 2;  
 /db\_xref="taxon:9606"  
 /clone\_lib="NC1\_USAP\_Skn3"  
 /lab\_host="DH10B (TI phage-resistant)"  
 /note="Morgan; skin; Vector: pCMV-SK4+6, Site: 1, NciI;  
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
 Average insert size 1.5kb. Library constructed by Life  
 Technologies. Note: this is a NC1\_USAP Library."

QY 1 GTGCTG 60  
 Db 1 GTGCTG 60  
 QY 61 TCTTCCATCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATAT 120  
 Db 61 TCTTCCATCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATATCAATAT 120  
 QY 121 CAAAG 180  
 Db 121 CAAAG 180  
 QY 181 TTTTCTGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 240  
 Db 181 TTTTCTGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 240  
 QY 241 AAG 300  
 Db 241 AAG 300  
 QY 301 TATAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
 Db 301 TATAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
 QY 361 CATCTGTTGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGAT 420  
 Db 361 CATCTGTTGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGAT 420  
 QY 421 CCGAG 480  
 Db 421 CCGAG 480  
 QY 481 TAGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGAT 540  
 Db 481 TAGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGAT 540  
 QY 541 GTATATCAAG 600  
 Db 541 GTATATCAAG 600  
 QY 601 AGTGGAG 660  
 Db 601 AGTGGAG 660  
 QY 661 GGTGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 718  
 Db 661 GGTGATATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAG 718  
 QY 719 CCAAG 777  
 Db 719 CCAAG 777  
 QY 779 GAAATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 837  
 Db 779 GAAATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 837  
 QY 839 GAAATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 891  
 Db 839 GAAATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATATGAGATAT 891

RESULT 4







```

Db 361 GACATGAGTTCCTGGAGAGTCAATGATTTCAAGAAATCTAAACATTTGATCGGGTGGAT 420
Qy 433 ATTTCTCCATTTTATATATATATATATATATATATATATATATATATATATATATAT 489
Db 421 ATTTCTCCATTTTATATATATATATATATATATATATATATATATATATATATATAT 480
Qy 429 TTTTCTCCATTTTATATATATATATATATATATATATATATATATATATATATATAT 549
Db 481 CGTTTCTCCATTTTATATATATATATATATATATATATATATATATATATATATAT 540
Qy 559 GAGTCTTTTATATATATATATATATATATATATATATATATATATATATATATATAT 609
Db 541 GAGTCTTTTATATATATATATATATATATATATATATATATATATATATATATATAT 600
Qy 615 TTGTAATATATATATATATATATATATATATATATATATATATATATATATATATAT 669
Db 601 GTTGAATATATATATATATATATATATATATATATATATATATATATATATATATAT 660
Qy 673 TTGTAATATATATATATATATATATATATATATATATATATATATATATATATAT 729
Db 661 GTTGAATATATATATATATATATATATATATATATATATATATATATATATATATAT 720
Qy 730 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
Db 721 ACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
Qy 790 CAA 732
Db 780 AAA 782

RESULT 7
LOCUS BG697846
DEFINITION BG697846 763 bp mRNA linear EST 07-MAY-2001
ACCESSION BG697846
VERSION BG697846.1 GI:13964518
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
REFERENCE 1 (bases 1 to 763)
AUTHORS NIH MGC project (http://mgi.scripps.edu/).
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Email: gregor.fremail@nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM10701 row: 6 column: 13
High quality sequence stop: 760.
FEATURES
Source location/Qualifiers
1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:490452"
/lab_host="NCI CGAP skn3"
/note="cloned from skin, Vector: pCMV-Sp65p; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 199 a 174 c 187 g 203 t
ORIGIN
Query Match 55.3%; Score 751; DB 12; Length 763;

```

```

Best Local Similarity 99.9%; Pred. No. 1.5e-123;
Matches 762; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 GTGGTGGTCTGCTGAGAGTCAATGATTTCAAGAAATCTAAACATTTGATCGGGTGGAT 60
Db 1 GTGGTGGTCTGCTGAGAGTCAATGATTTCAAGAAATCTAAACATTTGATCGGGTGGAT 60
Qy 61 TCTTCCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 120
Db 61 TCTTCCATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 120
Qy 121 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Qy 181 TTTCTGTAATATATATATATATATATATATATATATATATATATATATATATATAT 240
Db 181 TTTCTGTAATATATATATATATATATATATATATATATATATATATATATATATAT 240
Qy 241 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 241 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Qy 301 TATAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 301 TATAACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Qy 361 CATCTGTTGACATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Db 361 CATCTGTTGACATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
Qy 421 CCGGGTGATAGTTTCTCCATTTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Db 421 CCGGGTGATAGTTTCTCCATTTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 480
Qy 481 TAGACATGCGGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTT 540
Db 481 TAGACATGCGGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTTGCAGAGCTT 540
Qy 541 GTATATCCAGAGCTTTTATATATATATATATATATATATATATATATATATATATATAT 600
Db 541 GTATATCCAGAGCTTTTATATATATATATATATATATATATATATATATATATATATAT 600
Qy 601 AGCGGGCTCTGATATATATATATATATATATATATATATATATATATATATATATATAT 660
Db 601 AGCGGGCTCTGATATATATATATATATATATATATATATATATATATATATATATATAT 660
Qy 661 GCTTCTCTATGTTGCTTGGGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 GCTTCTCTATGTTGCTTGGGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Qy 721 AAGCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 762
Db 721 AAGCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 762

RESULT 8
LOCUS AL572354
DEFINITION AL572354 LTI_NFL006_FL2 Homo sapiens cDNA clone (NM001007)B07.5
prime, mRNA sequence.
ACCESSION AL572354
VERSION AL572354.1 GI:12930539
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
REFERENCE 1 (bases 1 to 783)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

```



DB 121 CAAGGAGCAAAATAGCTTTCTTTGGCGGAATGCTTTAAATTCCTCCATATGGAGCGAAATAT 180  
QY 18 TTTCTCTGAATATGACAGAGAAATTTCTCTGTGAATCCACACAGAAAGCTTTCTCTGAAGTAGCA 240  
DB 181 TTTCTCTGAATATGACAGAGAAATTTCTCTGTGAATCCACACAGAAAGCTTTCTCTGAAGTAGCA 240  
QY 241 AAGGAATGAGGAGAT 300  
DB 241 AAGGAATGAGGAGAT 300  
QY 301 TATACACAGCTGTTCT 360  
DB 301 TATACACAGCTGTTCT 360  
QY 361 CATCT 420  
DB 361 CATCT 420  
QY 421 CCGGAGTATAGTTCT 480  
DB 421 CCGGAGTATAGTTCT 480  
QY 481 TAGCAGATGCGTTCT 540  
DB 481 TAGCAGATGCGTTCT 540  
QY 541 GTATATCCAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCTTTTAAATCTGACACATAT 600  
DB 541 GTATATCCAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCTTTTAAATCTGACACATAT 600  
QY 601 AGCGGAGCTGTTGATAATCAGGTGTATGTGCGACAGAGCTTTTAAATCTGACACATATATCTCATTT 660  
DB 601 AGCGGAGCTGTTGATAATCAGGTGTATGTGCGACAGAGCTTTTAAATCTGACACATATATCTCATTT 660  
QY 661 GCTCTCTATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
DB 661 GCTCTCTATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
QY 721 AAGGCTGGACAGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764  
DB 721 AAGGCTGGACAGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764

RESULT 10  
BM542387  
LOCUS  
DEFINITION: DME42387 1013 bp mRNA linear EST 20.FEB.2002  
AGENCY: NCBI, NIH, MGC, 70 Homo sapiens cDNA clone IMAGE:5539560  
ACCESSION  
BM542387  
VERSION  
1  
MISC\_FEATURES  
EST  
SOURCE  
human  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1. (bases 1 to 1013)  
NIH-MGC http://www.ncbi.nlm.nih.gov/.  
TITLE  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL  
Unpublished (1999)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cchapb-remail.nih.gov  
Tissue Procurement: ATCC/RCMP/ntp  
cdna Library Preparation: Life Technologies, Inc.  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LMNI)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone deposition: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNI at:  
http://image.llnl.gov  
Plate: LMNI2234 row: e column: 01  
High quality sequence stop: 603.  
Location/Qualifiers  
1..1013  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="IMAGE:5539560"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_hosts="DH10B (phage resistant)"  
/notes="Organ: skin; Vector: pCMV-Sport6, Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 279 a 235 c 242 g 257 t  
OPIGIN

Query Match 54.3%; Score 738.2; DB 13; Length 1013;  
Best Local Similarity 96.7%; Pr-J No. 2.5e-121;  
Matches 766; Conservative 0; Mismatches 23; Indels 4; Gaps 3;  
QY 1 GTGGTGTCTGTCTGCAGAGTCATGACCTCTTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
DB 1 GTGGTGTCTGTCTGCAGAGTCATGACCTCTTTCCGCTTGGCCCTCAICGAGCTTCAGATT 78  
QY 61 TCTCCATCAAAATCAGATAAAGCTCACT 120  
DB 61 TCTCCATCAAAATCAGATAAAGCTCACT 138  
QY 79 TCTTCATCAAAATCAGATAAAGCTCACT 180  
DB 79 TCTTCATCAAAATCAGATAAAGCTCACT 198  
QY 101 CAAGGAGCAAAATAGCTTTCTTTGCGGAATGCTTTTAAATCTGACACATATATCTCATTTGAGAGCT 240  
DB 101 CAAGGAGCAAAATAGCTTTCTTTGCGGAATGCTTTTAAATCTGACACATATATCTCATTTGAGAGCT 258  
QY 181 TTTCTCTGAATATGACAGAGAAATTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
DB 181 TTTCTCTGAATATGACAGAGAAATTTCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318  
QY 241 AAGGAATGAGCAGATATATCTCATTTGAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCT 360  
DB 241 AAGGAATGAGCAGATATATCTCATTTGAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCT 378  
QY 301 TATACACAGCT 420  
DB 301 TATACACAGCT 438  
QY 361 CATCTCTGACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
DB 361 CATCTCTGACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 498  
QY 421 CCGGAGTATAGTTCT 540  
DB 421 CCGGAGTATAGTTCT 558  
QY 481 TAGCAGATGCGTTCT 600  
DB 481 TAGCAGATGCGTTCT 618  
QY 541 GTATATCCAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCTTTTAAATCTGACACATAT 660  
DB 541 GTATATCCAGAGCTTTTAAATCTGACACATATATCTCATTTGAGAGCTTTTAAATCTGACACATAT 678  
QY 601 AGCGGAGCTGTTGATAATCAGGTGTATGTGCGACAGAGCTTTTAAATCTGACACATATATCTCATTT 710  
DB 601 AGCGGAGCTGTTGATAATCAGGTGTATGTGCGACAGAGCTTTTAAATCTGACACATATATCTCATTT 738  
QY 661 GCTCTCTATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764  
DB 661 GCTCTCTATGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 776  
QY 720 CAAGGCTGGACAGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792  
DB 720 CAAGGCTGGACAGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800  
QY 777 TGAATATAGCGCAGCAAAATCCCCGTTTCTTTAGACA 809  
DB 777 TGAATATAGCGCAGCAAAATTTCCCCGTTTCTTTTAAA 831



```

/clone lib="UI-CF-EC1"
/tissue type="Lung"
/dev stage="Adult and Fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-EC1 is a normalized cDNA library containing the following tissue(s): Normal lung from adult and fetal day 64, day 87, week 19 and week 42. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is AAGTGCTTAC.
TAG LIB=UI-CF-EC1
TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371 and 380-383
TAG SEQ=AAGTGCTTAC"
BASE COUNT 193 a 159 g 230 t 1 others
ORIGIN

Query Match 53.6%; Score 728; DB 14; Length 760;
Best Local Similarity 99.0%; Pred. No. 1.7e-119;
Matches 754; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 199 ATATGACAGAGAAATTTTGTTGTAATCCACACAGAGAGCTTTCTGAAGTAGCAAGGAATG 248
Db 760 ATATGACAGAGAAATTTTGTTGTAATCCACACAGAGAGCTTTCTGAAGTAGCAAGGAATG 703

QY 249 CAGCATATATTCATTTGAGGTTTATTTCTGAAAGAGATGCTGGGAAATATATAACAC 308
Db 702 CAGCATATATTCATTTGAGGCTTATCTCTGAGAGAGATGCTGGGAAATATATAACAC 643

QY 309 CTGTCTCTGTTTGGGCTGATGGAGCTTTACTAGCAAGATATAGAAAGATCCATCTGTT 368
Db 642 CTGTCTCTGTTTGGGCTGATGGAGCTTTACTAGCAAGATATAGAAAGATCCATCTGTT 583

QY 369 TGACATATGATTTCTTGAGAAATTTACATTTCAAGAAATCTAAACATTTAGTCCGGGTGA 428
Db 532 TGACATATGATTTCTTGAGAAATTTACATTTCAAGAAATCTAAACATTTAGTCCGGGTGA 523

QY 429 TAGTTCTCCACATTTGATCTCTTATTCAGAGAGTGGTCTGGGCTCTGCTAGACAT 488
Db 522 TAGTTCTCCACATTTGATCTCTTATTCAGAGAGTGGTCTGGGCTCTGCTAGACAT 463

QY 489 GGGGTTTSCAGAGCTTGGCAAAATCTACCCACAGAGAGCTGCCAGCTCTTGTATATCC 548
Db 462 GGGGTTTSCAGAGCTTGGCAAAATCTACCCACAGAGAGCTGCCAGCTCTTGTATATCC 403

QY 549 AGGAGTTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 608
Db 402 AGGAGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 343

QY 609 TCTTATATCATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCT 668
Db 342 TCTTATATCATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCT 283

QY 669 TCTTATATCATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCT 728
Db 282 TCTTATATCATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCT 223

QY 729 CACAGAGAGAGTAAATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTAT 788
Db 222 CACAGAGAGAGTAAATGCTGTATATGCTGTATATGCTGTATATGCTGTATATGCTGTAT 163

QY 789 GCAATATCCCGGTTTTTAGACACAGAGGATCAGACCTCTATGCTGTAGATGAAAAGCC 848
Db 162 GCAATATCCCGGTTTTTAGACACAGAGGATCAGACCTCTATGCTGTAGATGAAAAGCC 103

```

```

QY 849 CTTAAAGTTTATGTTTCTTAATGTTCTACAGAGATAGGAGATATGATTTCTACACATAATCA 908
Db 102 CTTAAAGTTTATGTTTCTTAATGTTCTACAGAGATAGGAGATATGATTTCTACACATAATCA 43

QY 909 ACTCCCT-ATTAATTTCTTTAATGAGAGAAAAAAATTTTAAAA 949
Db 42 ACTCCCTNATTAAATTTCTTTAATGAGAGAAAAAAATTTTAAAA 1

RESULT 13
BM474944
LOCUS BM474944
DEFINITION AGENCOURT 647606 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:552749
5', mRNA sequence.
ACCESSION BM474944
VERSION RM474944 1 GI:18523986
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 987)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at.
http://image.llnl.gov
Plate: LLAM12292 Row: k Column: 05
High quality sequence stop: 685.
Location/Qualifiers
1. 987
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="IMAGE:552748"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI. Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH MGC Library."
BASE COUNT 258 a 240 c 241 g 247 t 1 others
ORIGIN

Query Match 52.5%; Score 714; DB 13; Length 987;
Best Local Similarity 99.6%; Pred. No. 4.8e-117;
Matches 747; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 1 GTGGTGCTTGTCTGAGAGTCATGAGCTTTTCCGCTTTCGCTTTCGCTTTCGCTTTCGCT 60
Db 48 GTGGTGCTTGTCTGAGAGTCATGAGCTTTTCCGCTTTCGCTTTCGCTTTCGCTTTCGCT 107

QY 61 TCTTCCATCAATTCAGATAAATCTGCTGCTTCTAGCTTCATCGGAGAGAGCAAGC 120
Db 108 TCTTCCATCAATTCAGATAAATCTGCTGCTTCTAGCTTCATCGGAGAGAGCAAGC 167

QY 121 CAAGAGAGCCAAATAGTTCTTCTGCGGAAATCTTAAATTCATATATATATATATATAT 180
Db 168 CAAGAGAGCCAAATAGTTCTTCTGCGGAAATCTTAAATTCATATATATATATATATAT 227

QY 181 TTTCTCTGATATGAGAGAAAAATTTCTGCTGATATCATACATCAAGAGCTTTCTTAAATAGCA 248
Db 228 TTTCTCTGATATGAGAGAAAAATTTCTGCTGATATCATACATCAAGAGCTTTCTTAAATAGCA 287

```



Source	Operation	Accession	Sequence
human.		480	AATCTACGACAGAGAGCTGCCAGCTCTGGTATATCCAGAGAGCTTTTAACTCTACCTAC 533
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		QY	750 TGGACACAGCCATTGGGAGTTTACTTTCAGCGAAGCGGGGTTTGTATATATATATATATAT 629
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo			
1. (bases 1 to 919)		Db	540 TGGACACAGCCATTGGGAGTTTACTTTCAGCGAAGCGGGGTTTGTATATATATATATAT 599
NIH-MGC <a href="http://mhc.nci.nih.gov/">http://mhc.nci.nih.gov/</a> .			
National Institutes of Health, Mammalian Gene Collection (MGC)		QY	630 GGCACACAGCCTCTCTCTGCCGGGATGACAAAGCCCTCCTATGCTTGGCTGGGACACAGATAC 689
Unpublished (1299)			
Contact: Robert Strausberg, Ph.D.		Db	600 GGCACACAGCCTCTCTCTGCCGGGATGACAAAGCCCTCCTATGCTTGGCTGGGACACAGATAC 659
Email: <a href="mailto:cgabbs@mail.nih.gov">cgabbs@mail.nih.gov</a>			
Tissue Procurement: ATCC		QY	690 CTGTGTGAACCTTTGGGGGAGGTTCTTAGCCAAAGCTGGG--CAACAGAAAGATATATG 745
cDNA Library Preparation: Life Technologies, Inc.			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		Db	660 CTGTGTGAACCTTTGGGGGAGGTTCTTAGCCAAAGCTGGG--CAACAGAAAGATATATG 719
DNA Sequencing by: Incyte Genomics, Inc.			
Clone description: MGC clone distribution information can be		QY	746 TGTATTCAGACATAGACCTTGAAAGAGCTGGCTGAA 780
found through the I.M.A.G.E. Consortium/LLNL at:		Db	720 TGTATTCAGACATAGA-CTGAAGAAAGCTGGGTGAA 753

Search completed: January 31, 2003, 11:21:20  
Job time : 2229 secs

Search completed: January 31, 2003, 11:21:20  
Job time : 2229 secs





QY 581 ATGGGAGTTACTTCCAGGAGCGCGGCTGTGATAATCAGGTGTATGTGGCCACACAGCTT 640  
 DB 785 ACTGGAGAGTGTGCTGCGGCGGCTGTGATAATCAGGTGTATGTGGCCACACAGCTT 644  
 QY 641 CTCCTGCGCGGATGACAAAGCTCTCTATTTTGGTGGGAAATAGAGGAGGAGGAGGAG 644  
 DB 845 CACAGTGTGGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 701 CTGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 DB 905 CTGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 761 ACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 DB 965 ACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 821 ACCTCTATG 829  
 DB 1025 ACCTCTATG 1033

RESULT 3  
 US-09-774-639-92  
 ; Sequence 92, Application US/09774639  
 ; Publication No. US20030003555A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: 90 Human Secreted Proteins  
 ; FILE REFERENCE: P2013P1  
 ; CURRENT APPLICATION NUMBER: US/09/774,639  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/214,112  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04  
 ; NUMBER OF SEQ ID NOS: 371  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 92  
 ; LENGTH: 1203  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (1165)  
 ; OTHER INFORMATION: n equals a.t.g. or c  
 US-09-774-639-92

Query Match 7.1%, Score 97; DB 9; Length 1203;  
 Best Local Similarity 50.9%; Pred No 9.7e-14;  
 Matches 256; Conservative 0; Mismatches 245; Indels 2; Gaps 1;

QY 641 ATGGGAGTTACTTCCAGGAGCGCGGCTGTGATAATCAGGTGTATGTGGCCACACAGCTT 640  
 DB 785 ACTGGAGAGTGTGCTGCGGCGGCTGTGATAATCAGGTGTATGTGGCCACACAGCTT 644  
 QY 641 CTCCTGCGCGGATGACAAAGCTCTCTATTTTGGTGGGAAATAGAGGAGGAGGAGGAGGAG 644  
 DB 845 CACAGTGTGGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 701 CTGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 DB 905 CTGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 761 ACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 DB 965 ACTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 644  
 QY 821 ACCTCTATG 829  
 DB 1025 ACCTCTATG 1033

RESULT 3  
 US-09-774-639-92  
 ; Sequence 92, Application US/09774639  
 ; Publication No. US20030003555A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: 90 Human Secreted Proteins  
 ; FILE REFERENCE: P2013P1  
 ; CURRENT APPLICATION NUMBER: US/09/774,639  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/214,112  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-04  
 ; NUMBER OF SEQ ID NOS: 371  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 92  
 ; LENGTH: 1203  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (1165)  
 ; OTHER INFORMATION: n equals a.t.g. or c  
 US-09-774-639-92

Query Match 7.1%, Score 97; DB 9; Length 1203;  
 Best Local Similarity 50.9%; Pred No 9.7e-14;  
 Matches 256; Conservative 0; Mismatches 245; Indels 2; Gaps 1;

QY 329 ATGGAGCTTTACTAGCAAGATATAGAAAGATCCATCTGTTGACATTTGATGTTCTCGGAA 388  
 DB 96 AAGGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388  
 QY 389 AAATTTACATTTCAAGATCTA--AAACATTTAGTCCGGGTGATAGTTTCTCCACATTGA 446  
 DB 156 AGGGCTATGTTGTGAAGCAACTCTACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 215  
 QY 447 TACTCTTACTGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 506  
 DB 216 CACACACAGAGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 275  
 QY 507 ACAAATCTACGACAGAGAGGCTGCCAGCTGTTGGTATATATCCAGGAGCTTTTAACTGAC 566  
 DB 276 TTTGGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 335  
 QY 567 TACTGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 626  
 DB 336 TACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 395  
 QY 627 TTTGGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 686  
 DB 396 TGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 455

QY 233 GAGGATGCTGGAAGATTTATATACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 342  
 DB 485 GAGGATGCTGGAAGATTTATATACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 544  
 QY 343 GCAGATGCTGGAAGATTTATATACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402  
 DB 545 GCT 604  
 QY 403 GAATCTA--AAACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 460  
 DB 605 AAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 664  
 QY 461 GAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 520  
 DB 665 AGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 724  
 QY 521 AGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 580  
 DB 725 AAGCTGGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784







QY 51 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 DL 51 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 QY 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 DB 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 QY 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 DB 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 QY 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276  
 DB 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276

RESULT 2  
 Q9JHW2 PRELIMINARY, PRT; 276 AA.  
 AC Q9JHW2  
 DT 01-MAR-2002 (Tremblrel. 20, created)  
 DT 01-MAR-2002 (Tremblrel. 20, last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, last annotation update)  
 DE Nit protein 2.  
 OS Homo sapiens (Human)  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 PC TISSUE=TESTIS;  
 RA Strausberg P.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 CE FMT: 00000000, AAH000001;  
 DE Interpro: IPR003010, Nitase/Chydase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 SC SEQUENCE 276 AA; 36580 MW; 10AA/97601E/7BR4C CPC64;

Query March 99.78; Score 1434; DB 4; Length 276;  
 Best Local Similarity 99.6%; Pred. No. 3.8e-124;  
 Matches 275; Conservative 0; Mismatches 1; Indels 0;  
 Gaps 0;  
 QY 1 MTSFLALIQICISIKSDNVTRACSFIREATOCAGKIVSLPECNSPYGAKYFPEYAEK 60  
 DB 1 MTSFLALIQICISIKSDNVTRACSFIREATOCAGKIVSLPECNSPYGAKYFPEYAEK 60  
 QY 61 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 DB 61 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 QY 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 DB 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 QY 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 DB 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 QY 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276  
 DB 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276

RESULT 3  
 Q9JHW2 PRELIMINARY, PRT; 276 AA.  
 AC Q9JHW2  
 DT 01-OCT-2000 (Tremblrel. 15, created)  
 DT 01-OCT-2000 (Tremblrel. 15, last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, last annotation update)  
 DE Nit protein 2 (1190017B10pik protein) (PIKEN cDNA 1190017B10)

DE gene).  
 GN NIT 09 1190017B10pik  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 PC PubMed=1095938;  
 RA Pace H.C., Hodawadekar S.C., Draganescu A., Huang J., Bieganski P.,  
 RA Pekarsky Y., Croce C.M., Brenner C.;  
 RT "Crystal structure of the worm NitPhit Rosetta Stone protein reveals a  
 RT Nit tetramer binding two Phit dimers";  
 RL Curr. Biol. 10:907-917(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 PC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa Y., Iwata M., Nishi Y., Fiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Batsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Ronald M.F.,  
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons F., Marchionni L., Mashima Y., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Shioh K. F.,  
 RA Suzuki H., Toyooka K., Wang Y.H., Weir G., Whitaker C., Wilming L.,  
 RA Wyshaw-Bolis A., Yoshida K., Yasogawa Y., Kawaji H., Kohsuke S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 PC TISSUE=MAMMARY TUMOR;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR ENBL, AC284573; AAP07102 1;  
 DR ENBL, AC284573; BAB23354 1;  
 DR ENBL, AC284573; BAB23354 1;  
 DR ENBL, BC020553; AAB20153 1;  
 DR MGD; MGI:1913477; 1190017B19Rik.  
 DR Interpro: IPR003010, Nitase/Chydase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 SC SEQUENCE 276 AA; 36580 MW; 74CFC449793D6D CPC64;

Query March 91.58; Score 1322; DB 11; Length 276;  
 Best Local Similarity 89.58%; Pred. No. 8.3e-114;  
 Matches 247; Conservative 17; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MTSFLALIQICISIKSDNVTRACSFIREATOCAGKIVSLPECNSPYGAKYFPEYAEK 60  
 DB 1 MTSFLALIQICISIKSDNVTRACSFIREATOCAGKIVSLPECNSPYGAKYFPEYAEK 60  
 QY 61 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 DB 61 IPGSESTOKLSEVAKESYVILIGSIFEDAGKLNITCAVEGPDGTLAKYKIHLPDIDV 120  
 QY 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 DB 121 PKKITFOESKTLSPGDSFSTFTTTCYRVLGICYNMRFAELAQIYAQPCQLLVPGAFN 180  
 QY 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 DB 181 LTGTGAHWELLQSRSAVNCVAVATASPARDDKASVAVAGHSTVNVFWGEVLAKAGTEEA 240  
 QY 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276  
 DB 241 IVYSIDILKZLAPICQIVFPFPPSTLYAVEMKPP 276

Db 241 LLYSDYDYLFLAEIPKQIPILPAPPADLYTVESKYP 276

## RESULT 4

Q8D0S1 ID Q8D0S1 PRELIMINARY FFT 131 AA.  
 AC Q8D0S1  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DE 1190017B19PIK  
 GN 1190017B19PIK  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 PP SEQUENCE FROM N.A.  
 RC STRAIN: C57BL/6J, TISSUE: EMERYO,  
 RX MEDLINE=21085660, PubMed=11217851;  
 PA Kawai J., Shinagawa A., Shibata Y., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa Y., Iwawa M., Nishi Y., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saio T., Okazaki Y., Gotoh T., Bono H., Kasukawa T., Saio R.,  
 Kadota Y., Marsuda H.A., Ashburner M., Baralov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsui S., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staibli F., Suzuki K., Tomita M., Wagner L., Washio T.,  
 PA Sakai K., Okido T., Furuno M., Anno H., Baldarelli P., Barsh G.,  
 Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann V., Hume D.A., Kariya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,  
 Nordone P., Ping B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schenbach C., Seva T., Shibata Y., Storch K.F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida Y., Hasegawa Y., Kawaji H., Kohsaki S.,  
 Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 FT  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF063604, BARC2884.1;  
 DR MGI: 1913477, 1190017B19PIK.  
 DR InterPro: IPR003010, Nucleo/hydase  
 DR Pfam: PF00795, CN hydrolase, 1.  
 SC SEQUENCE 131 AA, 2535 MW, 91741F4FE3CF40 CPOK4,

Query Match 74.8%, Score 1074, DB 11, Length 231,  
 Best Local Similarity 91.8%, Pred No. 4, 9e-21,  
 Matches 201, Conservative 11, Mismatches 7, Indels 0, Gaps 0.  
 1 MTSFFALQLQICSTKSRVITFCSTIFPAATGGAKIVSLPFCNIVGAKYFEVAEK 60  
 1 MSTTFALQLQICSTKSRVITFCSTIFPAATGGAKIVSLPFCNIVGAKYFEVAEK 61  
 61 IQGSECKYSEKAFKGVYIGGICSEEDACKYNTAVGSGNITLAKYFKHLEIDV 120  
 61 IDSECKYSEKAFKGVYIGGICSEEDACKYNTAVGSGNITLAKYFKHLEIDV 120  
 121 PGKTFKESFTTSPKSPSTFDTYTPVGLGICDMPFAELAQIYAQPGCLLVYPAEN 180  
 121 PGKTFKESFTTSPKSPSTFDTYTPVGLGICDMPFAELAQIYAQPGCLLVYPAEN 180  
 181 LTGPAHWELLPSPAVLNVVYVATASPAKDKASYVAV 219  
 181 LTGPAHWELLQAPAVDQNVVYVATASPAKDKASYVAV 219

## RESULT 5

Q8CTG9 ID Q8CTG9 PRELIMINARY FFT 189 AA.  
 AC Q8CTG9  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE 1190017B19PIK protein (fragment)  
 GN 1190017B19PIK  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 PP SEQUENCE FROM N.A.  
 RC STRAIN: C57BL/6J, TISSUE: EMERYO,  
 RX MEDLINE=21085660, PubMed=11217851;  
 PA Kawai J., Shinagawa A., Shibata Y., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Aizawa Y., Iwawa M., Nishi Y., Kiyosawa H., Kondo S., Yamanaka I.,  
 Saio T., Okazaki Y., Gotoh T., Bono H., Kasukawa T., Saio R.,  
 Kadota Y., Marsuda H.A., Ashburner M., Baralov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsui S., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staibli F., Suzuki K., Tomita M., Wagner L., Washio T.,  
 PA Sakai K., Okido T., Furuno M., Anno H., Baldarelli P., Barsh G.,  
 Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann V., Hume D.A., Kariya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,  
 Nordone P., Ping B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schenbach C., Seva T., Shibata Y., Storch K.F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida Y., Hasegawa Y., Kawaji H., Kohsaki S.,  
 Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 FT  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF063604, BARC2884.1;  
 DR MGI: 1913477, 1190017B19PIK.  
 DR InterPro: IPR003010, Nucleo/hydase  
 DR Pfam: PF00795, CN hydrolase, 1.  
 DR FSCITE, FSC0443, GATAGE\_TYPE\_II, ERYNOR\_1.  
 FT NON TER 1  
 SC SEQUENCE 189 AA, 2183 MW, 4F1D743F9C7D2A DCC4,

Query Match 65.0%, Score 934, DB 11, Length 189;  
 Best Local Similarity 91.8%, Pred No. 3, 1e-78;  
 Matches 172, Conservative 11, Mismatches 6, Indels 0, Gaps 0.  
 46 ECAGPQNTCAVEFPGILLARYPKHLEIDVFPITFEFTSECFITFTIYER 147  
 1 ECAGPQNTCAVEFPGILLARYPKHLEIDVFPITFEFTSECFITFTIYER 148  
 148 VGLGICDMPFAELAQIYAQPGCLLVYPAENITLAKYFKHLEIDV 197  
 61 VGLGICDMPFAELAQIYAQPGCLLVYPAENITLAKYFKHLEIDV 198  
 198 PAKDPAFASVAVSGTGVVHWDEVLAFAVTEALVYSSCLLPFLAEIRLIPVFFKPT 267  
 121 PAKDPAFASVAVSGTGVVHWDEVLAFAVTEALVYSSCLLPFLAEIRLIPVFFKPT 267  
 268 LYAVEMKPE 276  
 181 LYTVESKYP 189

## RESULT 6

Q8RUF8 ID Q8RUF8 PRELIMINARY FFT 169 AA.  
 AC Q8RUF8  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE ATSG12040/F14F18.210.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicotyledons;  
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsis.





RA Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X.,  
RA Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D.,  
RA Merkulov, G., Milichina, N.V., McBarty, C., Morris, J., Mostreffi, A.,  
RA Mount, S.M., Moy, M., Murphy, R., Murphy, L., Murry, D.M., Nelson, D.L.,  
RA Nelsen, D.R., Nelsen, K.A., Nelsen, E., Nelsen, J., Nelsen, J.M., Nelsen, J.M.,  
RA Palazzolo, M., Pirman, G.S., Pan, E., Pollard, J., Puri, V., Reese, M.G.,  
RA Reinert, K., Remington, K., Sanders, P.D., Scheeler, P., Smith, T.,  
RA Rhine, B.C., Siden-Kiamos, I., Simpson, P., Skupski, M.P., Smith, T.,  
RA Shue, E., Spradling, A.C., Stapleton, M., Strong, P., Sun, E.,  
RA Svirskas, P., Tector, C., Turner, P., Venter, E., Wang, A.H., Wang, X.,  
RA Wang, Z.-Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J.,  
RA Williams, S.M., Woodard, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A.,  
RA Ye, J., Ye, S.F., Zaveri, L., Zhan, M., Zhang, G., Zhao, C., Zheng, L.,  
RA Zheng, X.H., Zhang, S.H., Zhang, W., Zhao, X., Zhu, S., Zhu, X., Smith, H.G.,  
RA Gibbs, P.A., Myers, E.W., Putlin, G.M., Venter, J.C.,  
RA "The genome sequence of *Drosophila melanogaster*";  
RA Science 287:2185-2195(2000).  
RN [2].  
RN SEQUENCE FROM N.A.  
RP STRAIN=EPF019;  
RP Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,  
RP Champagne M., Chavez C., Dorsett V., Drosnek D., Farfan D., Frise E.,  
RP George P., Gonzalez M., Guzman H., Kronmiller P., Li, P., Liao G.,  
RP Miranda A., Mungall C.J., Nuncio J., Pachek J., Paragis V., Park S.,  
RP Patel S., Phoumenawong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
RP Ceiniker S.,  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RP FMBL, AE003692, AAF04370.1, ...  
RP EMBL: AY095190; AAM12283.1; ...  
RP FlyBase, FB00037687, CG8130.  
RP InterPro: IPR001110; Ntase/Chydase.  
RP Pfam: PF00795; CN\_hydrolase, 1.  
SQ SEQUENCE 283 AA; 31870 MW; 46CF7A83CFDD17E CRC64;  
Query Match 54.3%; Score 781.5; DB 5; Length 281;  
Best Local Similarity 55.1%; Pred. No. 6 co-64;  
Matches 152; Conservative 51; Mismatches 64; Indels 9; Gaps 5;  
QY 5 PLALDGLTSSIKSDNVTACSPFPAATCAK--IVSLDECRNSPVGKVFYVAKZIP 62  
DB 9 PLALLQKSKKQVAVQNAVTV--EAAVPEHPRLLTDECFNAPVTKVFRYSYETIP 67  
QY 63 -GSTGKI SEVAPESTVILKQK---BEDAKVNTCAVPGDGLLAKYKPIHLFDI 118  
DB 62 GAVTSQENLAPFHAVYVGGTPELGENDA--IYNTCTVWSGTDLVAKRKHLPDI 125  
QY 115 EYPSKITPESKTLSP--IGSFSTFQYCPVSLVGYMPFAELAQVYAPQ3CQLLVYPGA 178  
DB 126 DVKGGIPPESESTLSAGNFTINLVGHKLGIGITVDIPEEYAPLYNAGCEMIYPA 185  
QY 179 FNLTTGPAWELIOPSPAVNGLVYVATAPAPTPASVYVANGHSTVVPWSEVYAPASTE 238  
DB 184 FMTTTPHWEKLEPSPAFENL--EVVTTSPAFSTSAKYVAYHSMVYVNEWAFVQSAS 245  
QY 239 EALVTSICLFLKAEPEQIEVFEPGSSDLYAVEMK 274  
DB 245 SEIVAVTIFSEVEVAP--LWVLSPELSEYATEPY 291  
RESULT 9  
ID CQ7839 PRELIMINARY; FRT; 272 AA.  
AC CQ7839.  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE N carbamoyl-D-amino acid amidohydrolase.  
GN MTH1811.  
OS Methanobacterium thermoautotrophicum.  
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
OC Methanobacteriaceae; Methanobacterium.  
OX NCBI\_TaxID=187420;  
RN [1].

RP SEQUENCE FROM N.A.  
RP STRAIN=DELTA H;  
RP MEDLINE=9803714; PubMed 9371463.  
RA Smith D.P., Donner-Schramm L.A., Deloughery C., Lee H.-M., Rubels J.,  
RA Adrege P., Bashirzadeh P., Eickly C., Goff, S., Silber, V.,  
RA Harrison D., Hoang L., Keagle P., Lamm W., Portier P., Qiu D.,  
RA Stadler F., Vicore P., Wang Y., Wierschewski, Y., Gifford, P.,  
RA Juwani N., Caruso A., Bush D., Safer H., Parwell P., Prabhakar S.,  
RA McQuigall S., Shimer J., Goyal A., Pietrovski S., Church G.M.,  
RA Daniels C.J., Mao J.-I., Rice P., Neelting J., Peavey J.N.,  
RP "Complete genome sequence of *Methanobacterium thermoautotrophicum*  
RP deltaH: functional analysis and comparative genomics";  
RP J. Bacteriol. 179:7135-7151(1997).  
RP FMBL, AE000444, AAB96277.1, ...  
RP InterPro: IPR001010; Ntase/Chydase.  
RP InterPro: IPR001110; UPF0012.  
RP Pfam: PF00795; CN\_hydrolase, 1.  
RP ProSite: PS01227; UPF0012; 1.  
RP Hydrolase; Complete proteome.  
SQ SEQUENCE 272 AA; 30889 MW; EEP1AC0FFAPF27AB CRC64;  
Query Match 49.5%; Score 712; DB 17; Length 202;  
Best Local Similarity 48.7%; Pred. No. 1,6e-57;  
Matches 123; Conservative 55; Mismatches 9; Indels 0; Gaps 0;  
QY 5 PLALDGLTSSIKSDNVTACSPFPAATCAK--IVSLDECRNSPVGKVFYVAKZIP 64  
DB 2 PLALLQKSKKQVAVQNAVTV--EAAVPEHPRLLTDECFNAPVTKVFRYSYETIP 61  
QY 65 STQKLSVAKESIVLIGSSIFEEAGKLVNTDAVEFEGTLLAKYKPIHLFVW-SKI 124  
DB 62 SIVTVPSTAPFGLHVAAGSIFEEETFEIVNTSFVMDGNTIGFHKVHLFVINVGEI 121  
QY 125 TFQSKTSLSPGKSTFPTVYCFVGHVGHVYMPFAELAQVYAPQ3CQLLVYPGA 184  
DB 122 SFPESDSLIAGNSVTVITTCVGVGVGYCYVGFPELSEWALGAEVLIFQVFMFTG 181  
QY 185 PAHWELLOPSAVINLVYVATASAPLCPASVYVANGHSTVVPWSEVYAPASTE 244  
DB 182 PAHWELLVKSPALINQCTVAASTAPRHSASVYVANGHSTVVPWSEVYAPASTE 241  
QY 245 DIDLVKLAIEPQGVVPPQPSDLY 269  
DB 242 EIVSAVETTFEFLSEPRFEDV 256  
RESULT 10  
ID CQ76091 PRELIMINARY; FRT; 327 AA.  
AC CQ76091.  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Nitrilase homolog 1.  
GN NIT1.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1].  
RP SEQUENCE FROM N.A.  
RP MEDLINE=9837986; PubMed=9671749;  
RA Tokarsky Y., Carfigli, M., Gifford, J., Drack, T., Goff, Y.,  
RA Tillib S., Draganesu A., Brenner, P., Rothman J.H., Huebner K.,  
RA Buchberg A.W., Wang A., Brenner, C., Goff, C.M.,  
RP "Nitrilase and Flit homologs are encoded as fusion proteins in  
RP *Drosophila melanogaster* and *Caenorhabditis elegans*";  
RP Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).  
RP EMBL: AF069984; AAC39901.1; ...  
RP FMBL: AF069984; AAC39901.1; ...  
RP InterPro: IPR001010; Ntase/Chydase.  
RP InterPro: IPR001110; UPF0012.  
RP Pfam: PF00795; CN\_hydrolase, 1.

Fri Jan 31 13:48:43 2003

```

DR PROGITH, PS01227; UPF0012; 1
SQ SEQUENCE 327 AA, 3596 MW, 99F7899MBA627R1 CFC64;

Query Match
33.6%; Score 483; DB 11; Length 290;
Best Local Similarity 38.3%; Pred. No. 2.4e-36;
Matches 107; Conservative 54; Mismatches 106; Indels 14; Gaps 6;

QY 2 TSFELALQLQIQLSSSI--KSNVITPACNSFIEAATQANTVSLSECESEFYCAVYFPE 56
DB 1 TSFELALQLQIQLSSSI--KSNVITPACNSFIEAATQANTVSLSECESEFYCAVYFPE 56
QY 5 TSWELPLVAVCVSTSTPNKQENFKTCABELVQEAARLGAFLAFLPEAFD--FIARNPAETL 62
DB 5 TSWELPLVAVCVSTSTPNKQENFKTCABELVQEAARLGAFLAFLPEAFD--FIARNPAETL 62
QY 57 YAKIHLESTPLSEFAVEGSIYVIGSSIFE--EFAKUNTCATVFGDGTLLAPY 110
DB 57 YAKIHLESTPLSEFAVEGSIYVIGSSIFE--EFAKUNTCATVFGDGTLLAPY 110
QY 63 LLSEFENGLLQVQSCQAFENSIW--SAGHEPESGWEKGFVIRWHLIRFSSVAVY 122
DB 63 LLSEFENGLLQVQSCQAFENSIW--SAGHEPESGWEKGFVIRWHLIRFSSVAVY 122
QY 111 EKIHLEFLLVETKATIFQSEKTLSTGDSFS--TPTTPTV--PVNIGICTMSPFAELAQIYAOR 169
DB 111 EKIHLEFLLVETKATIFQSEKTLSTGDSFS--TPTTPTV--PVNIGICTMSPFAELAQIYAOR 169
QY 123 KATHLCDEVEIFSGGVRRESNTHFGTILEFTVPTACVGLAICYDMEFPELSLKLAAAG 182
DB 123 KATHLCDEVEIFSGGVRRESNTHFGTILEFTVPTACVGLAICYDMEFPELSLKLAAAG 182
QY 170 COLLVTPSAGRLITTPAWEHLLORSFAVTCVGVATATASAFETVACVAVANGHSTVVNMG 229
DB 170 COLLVTPSAGRLITTPAWEHLLORSFAVTCVGVATATASAFETVACVAVANGHSTVVNMG 229
QY 183 ABILTYPSAFESVTPFAHWEVILLRKAIESQCTIV--AAAGCGRHRETRATYTHGVVYVW 242
DB 183 ABILTYPSAFESVTPFAHWEVILLRKAIESQCTIV--AAAGCGRHRETRATYTHGVVYVW 242
QY 230 EYLAAGSTEEA--VWGDIDLPKLAETPQCTIPVFFPFSSDIY 269
DB 230 EYLAAGSTEEA--VWGDIDLPKLAETPQCTIPVFFPFSSDIY 269
QY 243 TVVARESEFSLCLAFILULHFLGCMFQHLFVFGHREPDLY 280
DB 243 TVVARESEFSLCLAFILULHFLGCMFQHLFVFGHREPDLY 280

RESULT 12
Q88526 PRELIMINARY; PRT; 323 AA.
ID Q88526 PRELIMINARY; PRT; 323 AA.
AC Q88526;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Nitritase homolog 1.
GN NIT1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RF MEDLINE-98317986; PubMed-9671749;
RX RA Tullis S., Campiglio M., Siprashvili Z., Druck T., Sedkov Y.,
RA Tullis S., Draganesu A., Wermuth P., Rothman J.H., Huebner K.,
RA Bachberg A.M., Mazo A., Brenner C., Croce C.M.;
KA "Nitritase and hlt homologs are encoded as fusion proteins in
RT drosophila melanogaster and caenorhabditis elegans.";
EMBL; AF069985; AAC40184.1; -
DR Proc. Natl. Acad. Sci. U.S.A. 95:8744-8749(1998).
DR MGD; MGI:1350916; Niti1.
DR InterPro; IPR003010; Ntlse/CNhydase
DR InterPro; IPR001110; UPF0012.
DR Pfam; PF00795; CN hydrolase; 1.
DR PROSITE; PS01227; UPF0012; 1.
DR FROSITE; PS01227; UPF0012; 1.
SQ SEQUENCE 323 AA, 35717 MW, 10151CEBL51DF2C1 CFC64;

Query Match
33.6%; Score 483; DB 11; Length 323;
Best Local Similarity 38.3%; Pred. No. 2.4e-36;
Matches 107; Conservative 54; Mismatches 105; Indels 14; Gaps 6;

QY 2 TSFELALQLQIQLSSSI--KSNVITPACNSFIEAATQANTVSLSECESEFYCAVYFPE 56
DB 2 TSFELALQLQIQLSSSI--KSNVITPACNSFIEAATQANTVSLSECESEFYCAVYFPE 56
QY 5 TSWELPLVAVCVSTSTPNKQENFKTCABELVQEAARLGAFLAFLPEAFD--FIARNPAETL 95
DB 5 TSWELPLVAVCVSTSTPNKQENFKTCABELVQEAARLGAFLAFLPEAFD--FIARNPAETL 95
QY 57 YAKIHLESTPLSEFAVEGSIYVIGSSIFE--EFAKUNTCATVFGDGTLLAPY 110
DB 57 YAKIHLESTPLSEFAVEGSIYVIGSSIFE--EFAKUNTCATVFGDGTLLAPY 110
QY 63 LLSEFENGLLQVQSCQAFENSIW--SAGHEPESGWEKGFVIRWHLIRFSSVAVY 156
DB 63 LLSEFENGLLQVQSCQAFENSIW--SAGHEPESGWEKGFVIRWHLIRFSSVAVY 156
QY 111 EKIHLEFLLVETKATIFQSEKTLSTGDSFS--TPTTPTV--PVNIGICTMSPFAELAQIYAOR 169
DB 111 EKIHLEFLLVETKATIFQSEKTLSTGDSFS--TPTTPTV--PVNIGICTMSPFAELAQIYAOR 169
QY 123 KATHLCDEVEIFSGGVRRESNTHFGTILEFTVPTACVGLAICYDMEFPELSLKLAAAG 215
DB 123 KATHLCDEVEIFSGGVRRESNTHFGTILEFTVPTACVGLAICYDMEFPELSLKLAAAG 215
QY 170 COLLVTPSAGRLITTPAWEHLLORSFAVTCVGVATATASAFETVACVAVANGHSTVVNMG 229
DB 170 COLLVTPSAGRLITTPAWEHLLORSFAVTCVGVATATASAFETVACVAVANGHSTVVNMG 229

```



us-09-855-294b-1.rspt

Fri Jan 31 13:48:43 2003

QY 57 -VAKIPGESTKLGFAVARECCSIYLLIGGIFE...EDAGKLYMTCAVEGPGTLLAKY 110  
 LL 96 LLEEFINGDLGJYSZARECJWLEJGFHEFGQWECQKTYNCHVLNKGSGVASY 155  
 QY 111 RKIHLFDIVPGKIFPQESKTI SPDSFS--TFDTPYCRVGLGICYDMRFAELAQIYAQBG 169  
 DL 154 RKTLLCTVRFPGQWFESENYPFGCTI FFPVTFASKVGLATCYDMRPFELSLKLAQAG 215  
 QY 179 QGLLVYFGAFNLTTGPAHWELLGCRSAVTHQVYVATASPARCDKASYVAMGHSSTVVNFWG 229  
 DB 216 ABILTYSSAFSGVTGPAHWEVLLFAPAI ESQCQVIAAQQGRHHETASVGHSMVVDPMG 275  
 QY 230 EVLAKAGTEEAIVVSDIDYKFLAEIPQOIPVPFQKSDLY 269  
 DB 276 TVVARQSEGPGLAPITLHFLQCMFCHI PVFQHRPDLY 315

Search completed: January 31, 2003, 07:52:17  
 b time : 38 secs







Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLAST at <http://image.llnwd.net>.  
Series: IPAL Plate: 37 Row: 2  
This clone was selected for full length sequencing because it met the following selection criteria: matched mRNA a: 9910459.

FEATURES SOURCE

```
1. 965
/organism="Homo sapiens"
/db_xref="LocusID:56954"
/db_xref="Laxon:9606"
/clone="MGC:22234 IMAGE:"
/tissue_type="Testis, em
/clone_lib="NH MGC_61"
/lab_host="DH10B"
/notes/Vector: pNPF-LIB"
32. 862
```

32. 862

```

/codon_start=1
/product="Nit protein 2"
/protein_id="AAH20620.1"
/db_xref="GI:18088311"
/translation="MTSEFLALIC
FNSYGAKYRPEYAEKIFGESTQ
DGTLLAKYKTHLPDIDVPGKITE
LAQIYAQRGQQLLYVPGAFNLITG
WGRSTVTVNFWSEALAKAGTTEAFV
"
```

BASE POINT	285 a	214 c	220 g	246 t
ORIGIN				

Color: Match  
69 28. 5001 940 6. PB 3. Length 965;

Query Match 99.2%, 0; Mismatches 0; Gaps  
Best Local Similarity 99.1%, Pred. No. 36-224; Indels 0; Gaps  
Matches 946; Conservative 9;

1 GTGGTGTTGTGTGAGAGTCATGAGCTCTTGGCTTATCAAGCTTACGATT 60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

120

61 TCTTCCATCAAAATCAGATAACGTCACCTGGCTGTAGCTTCAATCGGGAAGTAAATG  
QY

121 CAAGGAGCCAAATAGTTTCTTTGCCGGAATGCTTTAATTCTCCATATGGAGCGAAATAT 180

1000

181 ~~TTTCCCTCATTATTCCTGTCGACAAAATTCTGTTGAATTCACACAGAAGCTTCTGAAGTAGCA~~ 240

Db 191 TTTCTGAATATGCAGAGAAAATCTGTGATCCACACAGAGAGATTTCTAGAGAA

[illegible]

55' AAGGAATGACCATATATTCATTGATGCTTATACCTGAAGAGGATGCTGGAAATTA 310

THE UNIVERSITY OF CHICAGO LIBRARY

\_\_\_\_\_

Db 311 TATAACACCTGTGCTGTGTTCATGTAACCTTTTACTAGCAAAGTATAGAAAGATC 319

100

Db 371 CATCTGTTGACATTGATGTTCTCTGGAAAAATTACATTTCAAGAACTCTAAAAACAATTGAGT 430

421 CCGGTGAAGTTCACCAATGGAACTGACAGAGTTGGTATT  
QY

490

QY  
481 TACGACATGGGTTCACAGCCCTGCACAATTCTACGCCAAGAAGAACTTATTT

[illegible]

6.2.1 GTATATCCAGGAGGCTTTAATCTGACCAATGGAAGAGGATGGAATTATTAAGGAA

QY 371 GIAHACCAOAGCTTTATTCASISSSSESSESSSS









saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
Direct Submission  
Unpublished  
2 (bases 1 to 152116)  
Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.  
Direct Submission  
Submitted (29-NOV-2001) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 152116)  
Kaul, R.F., Olson, M.V., Zhou, Y., James, P.A., Rouse, G., Wu, Z.,  
Saenphimmachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.  
Direct Submission  
Submitted (28-MAR-2002) Genome Center, University of Washington,  
Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 152116)  
This sequence version replaced gi|17149456.

```

----- Genome Center
Center: University of Washington, Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchgs@u.washington.edu
----- Project Information
Center project name: Chr-1
Center clone name: wpl-3a6114 (sc0322)

```

```
----- Summary statistics -----
Sequencing vector: plasmid; L087547; 100% of reads
Chemistry: Dye-terminator B1; 75% of reads
Chemistry: Dye-terminator B1g Cye; 23% of reads
Assembly program: Phrap; version 0.903119
Consensus quality: 152014 bases at least Q40
Consensus quality: 152110 bases at least Q30
Consensus quality: 152116 bases at least Q20
Insert size: 152116; sum-of-contigs
Quality coverage: 9.5x in Q20 bases; sum of contigs
```

Overlapping Sequences:  
5': Mapping in progress  
3': Mapping in progress

```

-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the RMAP assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality > 30); an attempt was made to resolve all regions with problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone, and the assembly was confirmed by restriction digest.

**Sequence Validation:** This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

	SeqDerMap	FingerPrint	SeqDerMap	FingerPrint
	-----	-----	-----	-----
	SeqDerMap	FingerPrint	SeqDerMap	FingerPrint
	-----	-----	-----	-----

[illegible][illegible]

842 AAAAGCCCAAAAGTTATGTTCTTAATCTGCACAGAAATAGACGATATGATTCTACAAC 901  
 34710 AAAAGTTCTAATACTTTATGTTCTCAAGCTGTCGCAAAACAAGAATACGATTCTGCAAC 34711  
 902 ATAAATCAAGCCCAATTAATAATCTTTTAATGAAGAAAAAAATTTAAAAAATAAAAAA 961  
 34710 ATAAATCAATCTTTATGTTCTTTAATGACAGATCCCTTTTAAATTCACCCCTTTTCCTCC 34651  
 962 AACCAATGTTTATATGACATGAGAAAATGTTATATATGTTGACATTTTCGACGGCAGATT 1021

b 34560 CAAGGAGCTCTCTATTAAAGATGATAATTCCTCAGCAAGCTGATATTTCCACACACCAAA 34591

b	34590	AAGTAGTAAAGAGG-----AGTGTGAGGCGAAGA- - - - -GCTGGCTCTGAAG	34547
y	34592	CTTCTCTATACATCAATATGCTTCTGAAACATATCAGATCTTGGTATCCT	1141
b	34546	TTCTCTCGGATTTTATGTTTCTCTACAGACAGCTTTTGTGAAATACCGATCTTGGTATCCT	34487
y	1142	GTGTATGGATTCACCTAATATAATAATATATTTGTGTCTGACCACTCTTAAAAAGCTTGCTGG	1201
	34486	GCTGATGATGATCACTAAATACAAATATATTTGTGTGTATGATCAAGAACCTCTTATCCACACTTGGTG	34427

[illegible]

y	1322	TCTTTGGGTGTT	1333
b	34316	TCTTTGGGTGTT	34305

RESULT 8	
ACCUSS	AC103591
DEFINITION:	Homo sapiens chromosome 1 clone pP1-36A114, complete sequence.
ACCESSION	AC103591
VERSION	AC103591.2
KEYWORDS	HTG

ORGANISM Homo sapiens  
SOURCE  
SOURCE: Homo sapiens.  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 152116)  
REFERENCE

11477	11355	3598	3555	8696	8686	-----	1367	1413	1155	1171	1885	2004
6382	6462	2067	2053	6	<800	-----	574	<800	232	<800	3807	3754
512	<800	7969	8134	1414	1398	-----	10692	10710	4622	4473	1410	1398
449	<800	3139	3108	3329	3346	-----	8515	8571	244	<800	1536	1530
5304	5224	1825	1797	246	<800	-----	2904	2939	-----	-----	8613	8686
3321	3266	5662	5556	1526	1530	-----	2600	2663	-----	9224	9170	
32	<800	10356	9973	7072	6848	-----	1196	1182	-----	731	757	
519	<800	3013	2974	8663	8686	-----	459	<800	-----	159	<800	
3097	3115	165	<800	818	757	-----	459	<800	-----	733	757	
3252	3266	5973	5817	1002	999	-----	2189	2175	-----	1002	999	
152	<800	7978	8134	1061	1063	-----	3769	3715	-----	13724	13809	
2462	2457	13969	13668	5458	5397	-----	221	<800	-----	6752	6848	
8777	8571	4492	4473	7892	7825	-----	7852	7810	-----	-----	-----	
936	948	449	<800	1828	1848	-----	6996	6930	-----	-----	-----	
37	<800	8602	9275	4274	4265	-----	-----	-----	-----	-----	-----	
1999	1960	685	<800	765	757	-----	1.152116	Location/Qualifiers	-----	-----	-----	
1094	1095	1216	1324	4790	4753	-----	/db xref="taxon.9606"	/organism="Homo sapiens"	-----	-----	-----	
4235	4200	2940	2974	6823	6848	-----	/chromosome="11"	/clone="RP11-386I14"	-----	-----	-----	
9300	9291	462	<800	6440	6369	-----	/clone_lib="RPC1 human BAC library 11"	/note="IS1 excised from between bases 133478 and 133479"	-----	-----	-----	
2707	2663	2948	2974	2945	2987	-----	133478..133479	-----	-----	-----	-----	
598	<800	1377	1324	1251	1230	-----	-----	-----	-----	-----	-----	
5657	5584	6	<800	1724	1732	-----	-----	-----	-----	-----	-----	
2007	2049	3543	3555	1565	1520	-----	-----	-----	-----	-----	-----	
2170	2175	1182	1171	3034	3139	-----	-----	-----	-----	-----	-----	
951	948	3453	3555	1051	1053	-----	-----	-----	-----	-----	-----	
1461	1413	1867	15879	1997	2004	-----	-----	-----	-----	-----	-----	
3689	3715	574	<800	1757	1732	-----	-----	-----	-----	-----	-----	
30	<800	3627	3555	2095	2120	-----	-----	-----	-----	-----	-----	
461	<800	120	<800	3662	3754	-----	-----	-----	-----	-----	-----	
2629	2663	7795	7710	225	<800	-----	-----	-----	-----	-----	-----	
7756	7810	2469	2519	5089	5000	-----	-----	-----	-----	-----	-----	
7500	7810	1351	1324	1274	1232	-----	-----	-----	-----	-----	-----	
9225	9291	1857	1797	2598	2682	-----	-----	-----	-----	-----	-----	
12	<800	5216	5130	3984	4009	-----	-----	-----	-----	-----	-----	
256	<800	8222	8134	2429	3105	-----	-----	-----	-----	-----	-----	
871	887	7068	6911	748	767	-----	-----	-----	-----	-----	-----	
470	<800	3668	3882	1911	1848	-----	-----	-----	-----	-----	-----	

Query Match

Best Local Similarity 78.8%; Pred. No. 3.7e-176;  
Matches 1050, Conservative 0, Mismatches 106, Indels 56, Gaps 11;

QY	2	TGGTGTTTGT
----	---	--

[illegible]

```

RESULT: 9
AC025249/c
LOCUS
DEFINITION
AC025249: 14,065 bp. ENA
HG 26-SEP 2000
Home sapiens clone P111-16CMT, WOPING DRAFT SEQUENCE, 19 unordered
Dices.

```



FEATURES		Location/Qualifiers	
source		1..1063	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		/map="C57BL/6J"	
		/clone="MGC:28240 IMAGE:3992930"	
		/tissue_type="Mammary tumor. WAP TGF alpha m=1; 7 months	
		old, gross tissue."	
		/clone_lib="NCI CGAP_Mam5"	
		/lab_host="DH10B"	
		/note="Vector: pCMV-SPORT6"	
		35..865	
		/codon_start=1	
		/product="RIKEN cDNA 1190017B19 gene"	
		/protein_id="AAH20153.1"	
		/db_xref="GI:18043304"	
		/db_xref="LOCUSID:66227"	
		/translation="MSTRLADIQLQVSSIKSDNITRACSLVPEAKOGANIVSLPEC FNSPYGTTFPDYAEKIKGESIKLSEVAKESSITLGGSIPELDAKILHTLCLH3P DGLILVHPKIHLPDIDVPGKTFPGSKTISPGSPSTPTPYCKVGLGICDMRPAE LAG1VAQSGCOLLVVGFANLTTFGARWELLCFAFVNTGVVATATFATTPAOWA WGHSTVDPWGQVLKAGTEETILYSDILKYLAEIPAGIPILKPPPPATITVFGKPP	
		"	
BASE COUNT		291 a	244 c 251 g 277 t
ORIGIN			
Query Match		50.1%; Score 681.4, DB 10, Length 1063;	
Best Local Similarity		85.3%; Pred. No. 1.8e-159;	
Matches		760; Conservative 0; Mismatches 131; Indels 0; Gaps 0;	
QY	1	GTGGTGTCTCTGCAGAGTCATGACCTCTTCCGGTTGGTGGTCTCATGTAAGCTTCAGATT	60
DB	14	GGGTGCACCGTGGAGAGGCAATGCTTACTTCCGGCTGGGCTTCATACACCTTCAATT	73
QY	61	TCTTCGATCAATCAGATAACGTCACCTGGCGCTTGTAGCTTCATGCGAGAGAGCAAG	120
DB	74	TCTTCGATCAATCAGATAACGTCACCTGGCGCTTGTAGCTTCATGCGAGAGAGCAAG	133
QY	121	CAAG	180
DB	134	CAAG	193
QY	181	TTTCTGATATGAG	240
DB	194	TTTCTGATATGAG	253
QY	241	AAGGAATGAG	300
DB	254	AAGGAG	313
QY	301	TATAACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT	360
DB	314	TATAACAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT	373
QY	361	CACTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	420
DB	374	CACTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	433
QY	421	CGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	480
DB	434	CGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	493
QY	481	TAGGACATGGGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	540
DB	494	TAGGACATGGGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	553
QY	541	GTATATCAG	600
DB	554	GTATATCAG	613
QY	601	AGCCAG	660
DB	614	AGCCAG	673

QY 902 ATAATCAACTCCCTATTAAATCTTTAAATGAAGAAAAAATTTAAAAAAGAAAAA 961

DB 106279 ATAATCAACTCCCTATTCTTTAAATGAAGATCCCTTTTAAATTTACCCCTTTCCCTTC 106211

QY 952 AATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021

DB 106210 CAAGAGAGCTCTCTTTAAGATGATAAATCTCAGAGAGCTGATTTTCCACACCAAT 106151

QY 1022 AATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081

DB 106150 AATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 106107

QY 1082 GTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1141

DB 106106 TTTCTTCCGATTTAAATTTCTCAGAGAGCTTTGTGAAAGTAACCUATCTTGTGTATCT 106047

QY 1142 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1201

DB 106047 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 105987

QY 1202 GAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1261

DB 105986 GAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 105932

QY 1262 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1321

DB 105931 TACCTT---AAAGTTCATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 105877

QY 1322 TCTTGTGATGAT 1333

DB 105877 TCTTGTGATGAT 105865

RESULT 10

LOCUS BC020153

DEFINITION Mus musculus, P15Y cDNA 1190017B19 gene, clone MGC 28240

IMAGE:3992930, mRNA, complete cds.

ACCESSION BC020153

VERSION BC020153.1 GI:18043303

KEYWORDS house mouse.

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE 1. (bases 1 to 1063)

Strausberg, P.

Direct Submission

Submitted (13 DEC 2002) to the National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A63, Bethesda, MD 20892-2590.

USA MGC T01-jan-02, WTS, bAF, /mgs.nci.nih.gov

Contract: MGC help desk

Email: cga@nci.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center.

Center code: BCM-HQSC

Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Guarantee: P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Fowles, J.P., Law-Mac, S., Martin, P.G., Murry, D.M., Richards, S., Gibbs, P.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IMAGE Plate: 36 Row: 3 Column: 19

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI: 1296354.





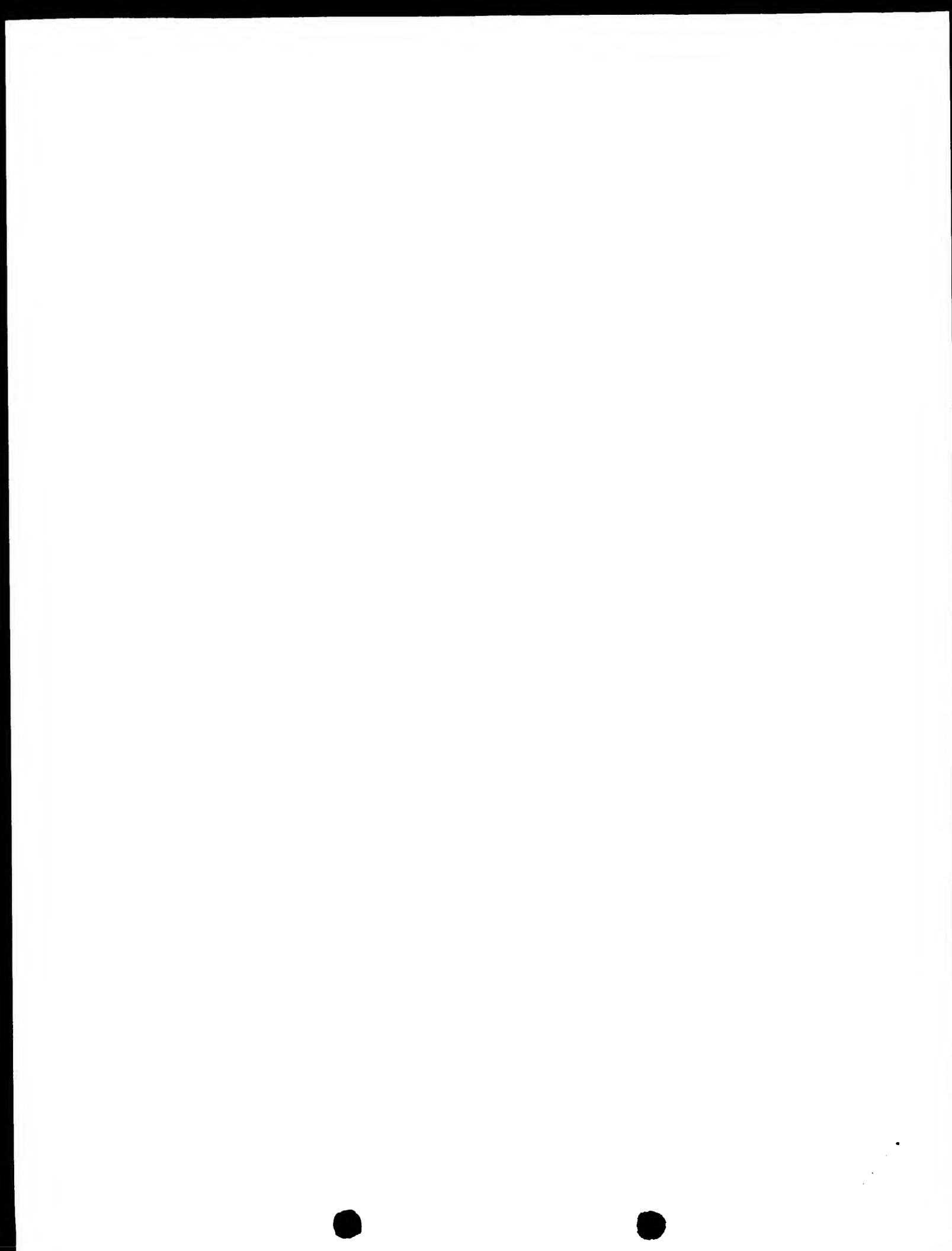
translation="MSTFRLALIQVSSSTKNDLTRACSLVREAAKQGANIVSLPECP  
FNPSGVYTFYPAEKIPGESTQKLSVEAKESSTLYGGST;PDEDKGLGHTICSVGPF  
DQSLVYKHKTHLEFIVPGSLTTPQESFTLYPGDSFTGTPGVYGLGH CYDMPFAF  
VLAIAQAQPGKTLIVTFAPNAPLQRAHWELIQRARAVDNOVYVATAS;PABEDKASIVA  
WGHSTWDPWGQVLTAKGTEETILYSOIDLAKLAEIRQQIPIILKQREADLYTVESKRP

[illegible]

RESULT 13  
AX473119

AX473119	AX473119	4043 bp	DNA	Linear	PAT 07 AUG 2000
LOCUS	Sequence 15 from Patent WO0194566.				
DEFINITION	AX473119				
ACCESSION	AX473119.1	GI:22207844			
VERSION					
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Cladialata, Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1				
TITLE	Poncho, G., Hilburn, E., Seoville, J., Turner, C.A., Friedrich, G., Abuin, A., Zambrowicz, P. and Sands, A.T.				
JOURNAL	Novel human enzymes and polynucleotides encoding the same				
FEATURES	Lexicon Genetics Incorporated (US)				
source	Location/Qualifiers				
	1..3093				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	829 a 677 c 704 g 883 t				
ORIGIN					
Query Match	37.4%; Score 508.2; DB 6; Length 3093;				
Best Local Similarity	98.5%; Pied. No. 3.5e-16;				
Matches	513; Conservative 0; Mismatches 8; Indels 0; Gaps 0;				
QY	442 TTGTGACATCGCTACTGCGACAGTGGTGTGGTATTTTATGACATATGTTGACAGAG 501				
DB	2168 TTGGAACACGCTACTGCGAGTGGTGTGGTATTTTATGACATATGTTGACAGAG 2227				
QY	502 GTTGCCAAATCTACGACACAGAGGTTGCGACCTGTGTGTATATCCAGGAGCTTTTAAT 561				
DB	2228 GTTGCCAAATCTACGACACAGAGGTTGCGACCTGTGTGTATATCCAGGAGCTTTTAAT 2287				
QY	562 CTACACCTGGACACAGCCGCTTTGGAGCTTATTCAGAGAGCCGGGTGTGATAATCAG 621				
DB	2288 CTACACCTGGACACAGCCGCTTTGGAGCTTATTCAGAGAGGCTGGTGTGATAATCAG 2347				
QY	622 GTGTATGTGGACAGAGCTTTGCTGTTGGGAGAGTTTATGCTAAAGCTGGCACAGAGAACA 741				
DB	2348 GTGTATGTGGACAGAGCTTTGCTGTTGGGAGAGTTTATGCTAAAGCTGGCACAGAGAACA 2407				
QY	682 CACAGACCGCTGGTGGACCTCTGGGGGAGAGTTTATGCTAAAGCTGGCACAGAGAACA 741				
DB	2408 CACAGACCGCTGGTGGACCTCTGGGGGAGAGTTTATGCTAAAGCTGGCACAGAGAACA 2467				
QY	742 ATGCTGTATTCAGACATAGACCTGAAGAAGCTGGTGAATATGCTGACAGAAATCCCTTT 801				
DB	2468 ATGCTGTATTCAGACATAGACCTGAAGAAGCTGGTGAATATGCTGACAGAAATCCCTTT 2527				
QY	802 TTGTAGACAGAGCGATCAGACCTCTATGCTGTGGAGATGAAAAGCTCTAAAGTTTATCT 861				
DB	2528 TTGTAGACAGAGCGATCAGACCTCTATGCTGTGGAGATGAAAAGCTCTAAAGTTTATCT 2587				
QY	862 TTCTTAATGTGTACAGAAATAGGACGATATGTTCTTACACATATATCAACTCCCTATTAAA 921				
DB	2588 TTCTTAATGTGTACAGAAATAGGACGATATGTTCTTACACATATATCAACTCCCTATTAAA 2647				
QY	922 TCTTTTAATGAGAAAAAAAATTTTAAAAAAAATAAAAAA 962				
DB	2648 TCTTTTAATGAGAAAAAAAATTTTAAAAAAAATAAAAAA 2688				
RESULT 14					
AX330407/c	AX330407	489 bp	DNA	Linear	PAT 07 JAN 2000
LOCUS	Sequence 916 from Patent WO0194629.				
DEFINITION	AX330407				
ACCESSION	AX330407.1	GI:18103385			
VERSION					
KEYWORDS					





GenCore version 5.1.3  
Copyright (c) 1993 GenCompen Ltd.

[illegible]

RUN ON: JANUARY 21, 2003, 07:42:23, Search time 37 seconds (without alignments)

Title: No. 99 950 004R ;

perfect score:

Sequences: MTSEFLYITLITLITLSEEN 1000 QIVAPPSLYAVEMKPT 174

Scoring table. *Prüfung*[illegible]

searched. 008470 eggs, 133,50620 residues

all number of five satisfactory been parameters

Minimum DB seq length: 0

[illegible]

Post-processing: Minimum Match of

Maximum Processing: Minimum: Match 100%  
Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1010002.\*

1:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA2001.DAT.*
23:	/SID527/gcd3a.a/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %		DB	ID	Description
		Match	Length			
1	1438	100.0	276	23	AA980984	Human hNit3 ase
2	1438	100.0	276	23	ABE76477	Human human nitric
3	1438	100.0	276	23	AAE15790	Human Nit2 protein
4	1384	96.2	285	23	ABE76483	Human human nitric
5	1332	91.9	276	23	AAE15791	Mouse Nit2 protein
6	781.5	54.3	283	23	ABE37310	Escherichia melanos
7	773	53.8	322	23	AAE15796	Schistosaccharomyce
8	740	51.5	173	23	APP76482	Human human nitric
9	700	48.7	159	21	AAE42987	Human GREX ORF2731
10	700	48.7	159	23	ABE76478	Mouse human nitric

## ALIGNMENTS

## RESULT 1

AA380984

ID AAB80984 standard, Protein, 276 AA.

AAC AAB80984:

10 JUN 2001

XX  
Human Rights - 3000  
DE

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

Waxville Hydro

OS Homo sapiens.

PM 127998-A.

27-DEC-2000.

XX  
DE 30-MAY-2000.XX  
30 MAY 2000 20000000112[illegible]

XX

XX

DR N. ESDE, AAF84

ed  
vxy

XX The present invention relates to human hNit3-ase (nitrlase) protein,  
CC which is expressed in Cushings adrenal gland tumour, and its coding  
CC sequence (AA04214 and AA05984). The present invention also relates to a  
CC preparation method of said protein and nucleic acid sequence, and a  
CC method of detecting human hNit3-ase nucleic acid sequence and polypeptide  
CC in sample.  
XX  
SQ Sequence 276 AA;

Query Match 100.0%; Score 1438; DB 22; Length 276;  
Best Local Similarity 100.0%; Pred. No. 1.5e-149;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

RESULT 2  
AAE15790  
ID AAE15790 standard; Protein; 276 AA.  
XX  
AC AAE15790;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human Nit2 protein.  
XX  
KW Human; Nit2; cytostatic; neuroprotective, cellular pathway; therapy;  
XX apoptosis; proliferative disorder; degenerative disease.  
XX Homo sapiens.

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

RESULT 3  
AAE15790  
ID AAE15790 standard; Protein; 276 AA.  
XX  
AC AAE15790;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human Nit2 protein.  
XX  
KW Human; Nit2; cytostatic; neuroprotective, cellular pathway; therapy;  
XX apoptosis; proliferative disorder; degenerative disease.  
XX Homo sapiens.

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

RESULT 4  
AAE15790  
ID AAE15790 standard; Protein; 276 AA.  
XX  
AC AAE15790;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human Nit2 protein.  
XX  
KW Human; Nit2; cytostatic; neuroprotective, cellular pathway; therapy;  
XX apoptosis; proliferative disorder; degenerative disease.  
XX Homo sapiens.

CC protein (NHP) that shares structural similarity with nitrlase  
CC enzymes from a wide variety of organisms. The invention provides  
CC nitrlase-like NHPs (see ABB76477-83) and polynucleotides (see  
CC ABN83808-15), host cell expression systems, fusion proteins,  
CC antibodies, antagonists, agonists, and other compounds that  
CC modulate NHP expression or activity. Those can be used to  
CC detect mutant or inappropriately expressed NHPs for the diagnosis  
CC of disease, or to screen for drugs effective in the treatment of  
CC perturbations of the normal function of an NHP. Engineered host  
CC cells and/or transgenic animals producing NHP can be used to  
CC identify compounds that bind the endogenous NHP receptor and to  
CC identify compounds that trigger NHP-mediated activities or  
CC pathways.  
XX  
SQ Sequence 276 AA;

Query Match 100.0%; Score 1438; DB 23; Length 276;  
Best Local Similarity 100.0%; Pred. No. 1.5e-149;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

RESULT 3  
AAE15790  
ID AAE15790 standard; Protein; 276 AA.  
XX  
AC AAE15790;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human Nit2 protein.  
XX  
KW Human; Nit2; cytostatic; neuroprotective, cellular pathway; therapy;  
XX apoptosis; proliferative disorder; degenerative disease.  
XX Homo sapiens.

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

RESULT 4  
AAE15790  
ID AAE15790 standard; Protein; 276 AA.  
XX  
AC AAE15790;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human Nit2 protein.  
XX  
KW Human; Nit2; cytostatic; neuroprotective, cellular pathway; therapy;  
XX apoptosis; proliferative disorder; degenerative disease.  
XX Homo sapiens.

QY 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
DB 1 MTSFRLALIQLOISSIKSDNVTPACSFIREAATOGAKIVSLPECFNSPYGAKYFPEYAEK 60  
QY 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
DB 61 IPGESTOKLSEVAKECSIYLGSIPEEDAGKLYNTCAVFGDGTLLAKYRKHLFDIDV 120  
QY 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
DB 121 PGKITFOESKTLSPGDSFSTFDTPCVGHGICVDMFPAELAQIYAQRGCOLLYVPGAEN 180  
QY 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
DB 181 LTGPAHWELLQSRVNDQVYVATASPARDDKASYVAMGHSTVVPWGEVLAKAGTEEA 240  
QY 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276  
DB 241 IVYSDIDLKLAE:EQQIPVPPQPPSELYAVEMFKP 276

FT Region 123..127  
FT /label= NS1\_strand  
FT Region 140..143  
FT /label= NS6b\_strand  
FT Region 146..151  
FT /label= NS7\_strand  
FT Region 153..164  
FT /label= NH3\_helix  
FT Region 171..178  
FT /label= NS8\_strand  
FT Region 182..199  
FT /label= NH4\_helix  
FT Region 201..213  
FT /label= NS9\_strand  
FT Region 216..226  
FT /label= NS10\_strand  
XX  
XX WO200187959-A2.  
PN  
22-NOV-2001.  
PF 15-MAY-2001; 2001WO-US15664.  
XX  
XX 16-MAY-2000; 2000US-204713F.  
XX  
XX (UYJE-) UNIV JEFFERSON THOMAS.  
XX  
XX Croce C, Brenner C, Pekarski Y;  
XX WPI; 2000-092984/11.  
DR N PSDB; AAD24256.  
XX

Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins,  
PT useful to find molecules that mimic or antagonize Phit interaction for  
PT the treatment of proliferative or degenerative diseases -  
XX  
XX Claim 5; Page 53-54; 51pp; English.

XX The invention relates to isolated nucleic acids comprising a fully  
XX defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
XX mouse Nit2 proteins. Nit and Phit proteins are encoded as fusion  
XX proteins in invertebrates and as separate polypeptides in vertebrates.  
XX Nit and Phit interact physically and functions in same cellular pathways.  
XX Molecules which bind Nit2 and mimic or antagonise Phit interaction are  
XX used to treat diseases in which activity of Nit2 protein is altered in  
XX a mammal. Phit mimics induce apoptosis and are particularly useful to  
XX treat proliferative disorders, whilst Phit antagonists promote cell  
XX proliferation and are particularly useful to treat degenerative disease.  
XX The present sequence is human Nit2 protein

XX SQ Sequence= 276 AA;

Query Match 100.0%; Score 1438; DB 23; Length 276.  
Best Local Similarity 100.0%; Pred No 1 52-149;  
Matches 276; Conservative 3; Mismatches 0; Indels 0; Gaps 0.

QY 1 MTSFPLALIGLQSSISFSTNVTATCSIPRAATGSAIVSLPEFNSFYGAKYFEVAEK 60  
DI 1 MTSFPLALIGLQSSISFSTNVTATCSIPRAATGSAIVSLPEFNSFYGAKYFEVAEK 60  
QY 61 IPRESTFPLSEVAPERSYILGDSIHFEQAFQVNTAVFPGTLLAKYFKHLEFDIV 120  
DI 61 IPRESTFPLSEVAPERSYILGDSIHFEQAFQVNTAVFPGTLLAKYFKHLEFDIV 120  
QY 121 PPKITPQESKTLSPGSPSTPDTPPVVSLGSIQYMPFAELAQIYAGGQGLLYVPSAFN 180  
DI 121 PPKITPQESKTLSPGSPSTPDTPPVVSLGSIQYMPFAELAQIYAGGQGLLYVPSAFN 180  
QY 181 LTTGPAHLLQSPAVNQQVYVATASPADDDKASYVAVGSHSTVYVNWGEVLAKAGTEEA 240  
DI 181 LTTGPAHLLQSPAVNQQVYVATASPADDDKASYVAVGSHSTVYVNWGEVLAKAGTEEA 240  
QY 241 IVYSEIDILKFLAEIPQIPVFRFFESLYAVENKXP 276

DB 241 IVYSEIDILKFLAEIPQIPVFRFFESLYAVENKXP 276

RESULT 4

ABB76483  
ID ABB76483 standard; Protein; 285 AA.

XX AC ABB76483;  
XX  
XX 23 SEP-2002 (first entry)  
XX Novel human nitrilase-like protein.  
XX Nitrilase; enzyme; human; NHP; gene therapy  
XX Homo sapiens.  
XX WO200194566 A2.  
XX  
XX 13-DEC-2001.  
XX  
XX 25-JAN-2001; 2001WO-US26647.  
XX  
XX 28-JAN-2000; 2000US-179000P.  
XX  
XX (LEXI-) LEXICON GENETICS INC.  
XX  
XX Donoho G, Hilbun E, Scoville J, Turner CA, Friedrich G, Abuin A,  
XX Zambrawicz E, Sands AT,  
XX WPI; 2002-519061/55.  
DR N-PSDB; ABH83814.

XX Novel polynucleotides encoding novel human proteins sharing sequence  
XX similarity with nitrilase proteins, useful for generating primers, and  
XX probes which are used in identification of targets for drug discovery  
XX  
XX Disclosure; Page 34-35; 36pp; English.  
XX  
XX The present sequence is the protein sequence of a novel human  
XX protein (NHP) that shares structural similarity with nitrilase  
XX enzymes from a wide variety of organisms. The invention provides  
XX nitrilase-like NHPs (see ABB76477-83) and polynucleotides (see  
XX ABN84808-16), host cell expression systems, fusion proteins,  
XX antibodies, antagonists, agonists, and other compounds that  
XX modulate NHP expression or activity. These can be used to  
XX detect mutant or inappropriately expressed NHPs for the diagnosis  
XX of disease, or to screen for drugs effective in the treatment of  
XX perturbations of the normal function of an NHP. Engineered host  
XX cells and/or transgenic animals producing NHP can be used to  
XX identify compounds that bind the endogenous NHP receptor and to  
XX identify compounds that trigger NHP-mediated activities or  
XX pathways.

XX SQ Sequence= 285 AA;

Query Match 99.8%; Score 1347; DB 23; Length 285;  
Best Local Similarity 97.1%; Pred No 1 42-143;  
Matches 285; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 MTSFPLALIGLQSSISFSTNVTATCSIPRAATGSAIVSLPEFNSFYGAKYFEVAEK 60  
DI 1 MTSFPLALIGLQSSISFSTNVTATCSIPRAATGSAIVSLPEFNSFYGAKYFEVAEK 60  
QY 61 IPRESTFPLSEVAPERSYILGDSIHFEQAFQVNTAVFPGTLLAKYFKHLEFDIV 120  
DI 61 IPRESTFPLSEVAPERSYILGDSIHFEQAFQVNTAVFPGTLLAKYFKHLEFDIV 120  
QY 121 PPKITPQESKTLSPGSPSTPDTPPVVSLGSIQYMPFAELAQIYAGGQGLLYVPSAFN 180  
DI 121 PPKITPQESKTLSPGSPSTPDTPPVVSLGSIQYMPFAELAQIYAGGQGLLYVPSAFN 180  
QY 181 LTTGPAHLLQSPAVNQQVYVATASPADDDKASYVAVGSHSTVYVNWGEVLAKAGTEEA 240  
DI 181 LTTGPAHLLQSPAVNQQVYVATASPADDDKASYVAVGSHSTVYVNWGEVLAKAGTEEA 240









CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematous, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 XX  
 SQ Sequence 159 AA;  
 Query Match 48.7%; Score 700; DB 21; Length 159;  
 Best Local Similarity 98.5%; Pred. No. 1.2e-68;  
 Matches 132; Conservative 1; Mismatches 1; Indels 6; Gaps 0;  
 QY 142 DTFFVYVGLGIDIMFPAELAQIYAFPGGQGLVYFSAFRLTTTFAHWELLQFSAVAVNQV 143  
 DL 25 ETATFVGLGIDIMFPAELAQIYAFPGGQGLVYFSAFRLTTTFAHWELLQFSAVAVNQV 84  
 QY 202 YVATASAFAPDPAFYVAMGSHSTVWVWGEVIAFAGTTEAIVYSDIDLFKLAETPQQIEVF 161  
 85 YVATASAFAPDPAFYVAMGSHSTVWVWGEVIAFAGTTEAIVYSDIDLFKLAETPQQIEVF 144  
 QY 262 PQPSPDIYAVEMKPP 276  
 145 PQPSPDIYAVEMKPP 159  
 Db 145 PQPSPDIYAVEMKPP 159  
 RESULT 10  
 ID ABB76478  
 AC ABB76478 standard; Protein; 159 AA.  
 AC ABB76478;  
 XX  
 DT 21-SEP-2002 (first entry)  
 XX  
 DE Novel human nitrilase-like protein  
 DE Nitrilase; enzyme; human; NHP; gene therapy.  
 KW  
 OS Homo sapiens.  
 XX  
 PN W0200194566.A2  
 PD 13 DEC-2001.  
 XX  
 PF 25 JAN-2001; 2001W019326687  
 XX  
 PP 28 JAN-2000; 2000US-179000P.  
 (LEXI-) LEXICON GENETICS INC.  
 XX  
 PI Donoho G, Hilbun B, Scoville J, Turner CA, Friedrich G, Abuin A;  
 PI Zambrowicz B, Sands AT;  
 XX  
 DP WPI; 2002-519061/55.  
 N-PSNB; ABN93809.  
 XX  
 FT Novel polynucleotides encoding novel human proteins sharing sequence  
 FT similarity with nitrilase proteins, useful for generating primers, and  
 FT probes which are used in identification of targets for drug discovery  
 PT  
 PT  
 PS Disclosure: Page 10-31, 36ff, English.  
 XX  
 CC The present sequence is the protein sequence of a novel human  
 CC protein (NHP) that shares structural similarity with nitrilase  
 CC enzymes from a wide variety of organisms. The invention provides  
 CC nitrilase-like NHPs (see ABB76477-83) and polynucleotides (see  
 CC ABN93809-15), host cell expression systems, fusion proteins,  
 CC antibodies, antagonists, agonists, and other compounds that  
 CC modulate NHP expression or activity. These can be used to  
 CC detect mutant or inappropriately expressed NHPs for the diagnosis  
 CC of disease, or to screen for drugs effective in the treatment of  
 CC perturbations of the normal function of an NHP. Engineered host

CC cells and/or transgenic animals producing NHP can be used to  
 CC identify compounds that bind the endogenous NHP receptor and to  
 CC identify compounds that trigger NHP-mediated activities or  
 CC pathways.  
 XX  
 SQ Sequence 159 AA;  
 Query Match 48.7%; Score 700; DB 23; Length 159;  
 Best Local Similarity 98.5%; Pred. No. 1.2e-68;  
 Matches 132; Conservative 1; Mismatches 1; Indels 6; Gaps 0;  
 QY 142 DTFFVYVGLGIDIMFPAELAQIYAFPGGQGLVYFSAFRLTTTFAHWELLQFSAVAVNQV 143  
 DL 25 ETATFVGLGIDIMFPAELAQIYAFPGGQGLVYFSAFRLTTTFAHWELLQFSAVAVNQV 84  
 QY 202 YVATASAFAPDPAFYVAMGSHSTVWVWGEVIAFAGTTEAIVYSDIDLFKLAETPQQIEVF 161  
 85 YVATASAFAPDPAFYVAMGSHSTVWVWGEVIAFAGTTEAIVYSDIDLFKLAETPQQIEVF 144  
 QY 262 PQPSPDIYAVEMKPP 276  
 145 PQPSPDIYAVEMKPP 159  
 Db 145 PQPSPDIYAVEMKPP 159  
 RESULT 11  
 ID AAE15794  
 AC AAE15794 standard; Protein; 291 AA.  
 AC AAE15794;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Saccharomyces cerevisiae Nlt3 protein.  
 DE Yeast; Nlt3; cytotactic; neuroprotective; cellular pathway; therapy;  
 KW apoptosis; proliferative disorder; degenerative disease.  
 XX  
 OS Saccharomyces cerevisiae.  
 XX  
 PH Key Location/Qualifiers  
 FT Region 9..19  
 FT /label= NS1\_strand  
 FT Region 24..42  
 FT /label= NHI\_helix  
 FT Region 47..52  
 FT /label= NS2\_strand  
 FT Region 64..88  
 FT /label= NH2\_helix  
 FT Region 92..103  
 FT /label= NS3\_strand  
 FT Region 105..117  
 FT /label= NS4\_strand  
 FT Region 122..127  
 FT /label= NS5a\_strand  
 FT Region 132..135  
 FT /label= NS5b\_strand  
 FT Region 139..143  
 FT /label= NS6a\_strand  
 FT Region 156..159  
 FT /label= NS6b\_strand  
 FT Region 162..168  
 FT /label= NS7\_strand  
 FT Region 175..185  
 FT /label= NH3\_helix  
 FT Region 187..194  
 FT /label= NS8\_strand  
 FT Region 198..215  
 FT /label= NH4\_helix  
 FT Region 217..229  
 FT /label= NS9\_strand  
 FT Region 232..242  
 FT /label= NS10\_strand  
 XX

WD2000187958-A2.  
 22-NOV-2001.  
 15-MAY-2001; 2001WC-US15664.  
 16-MAY-2000; 2000US-204/13P.  
 (UNVE-) UNIV JEFFERSON THOMAS.  
 Croce C, Brenner C, Pekarski Y;  
 WP1: 2002-092984/11.  
 Isolated cDNA encoding human, mouse, frog and yeast Nit2 proteins, for  
 useful to find molecules that mimic or antagonize Phit interaction for  
 the treatment of proliferative or degenerative diseases -  
 Claim 17; Page 56-57; 61pp; Eng.ish.  
 The invention relates to isolated nucleic acids comprising a fully  
 defined cDNA nucleotide sequence encoding human, Xenopus laevis and  
 mouse Nit2 proteins. Nit and Phit proteins are encoded as fusion  
 proteins in invertebrates and as separate polypeptides in vertebrates.  
 Nit and Phit interact physically and functions in same cellular pathways.  
 Molecules which bind Nit2 and mimic or antagonise Phit interaction are  
 used to treat diseases in which activity of Nit2 protein is altered in  
 a mammal. Phit mimics induce apoptosis and are particularly useful to  
 treat proliferative disorders, whilst Phit antagonists promote cell  
 proliferation and are particularly useful to treat degenerative disease.  
 The present sequence is yeast Nit3 protein.  
 Sequence 291 AA;  
 Query Match 48.3%, Score 294 e; DR 23, Length 291;  
 Best Local Similarity 49.8%; Pred. No. 1.3e-67;  
 Matches 137; Conservative 52; Mismatches 77; Indels 9; Gaps  
 4;  
 QY 5 RLALQLQISS-ISKDNVTACSPIRAATQ--CAKIVSLPCFNSPYGAKEYPEYAE-- 59  
 DB 12 KVALVQLSGSPKMANLCPAATFIRAMKQPDTKLVVLPCFNSPYSTQFRKYSEVI 71  
 QY 50 -KIPGESTQFLSEVAKECGIVLGGSIPEED -AGKLYNTCAVFGPGDTLLAKYFKIHL 115  
 DB 72 NPKPSTSVQFLNLANKFKIILVSGTTEIPDPKTKLYNTSLIFNEDGKLIDKRRKVL 131  
 QY 116 FDDVPGKIITFQSKTISPDGFSFTFTPTPYCPVGLGICVDWKFPAELADQIYAGKGCQLLVY 175  
 DB 132 FQVDIPRGISPHESSETISPGFESTIPDKYGRFWISLYDWFPEPELAMSARKGAFAFIY 191  
 176 PGAPNLITGFAHWELLORSRAVDNQVVATASPARDKAGYVANGHSITVVRPWGVLAKA 235  
 DB 192 PSAFTVTVGELJHLLAPSPAVDNCVVMLQSPARNLQSNHAYGHSIVVDPRKIVAE 251  
 QY 236 STEEAIVISDITFELAFETQGIIVFPCKRSDIYA 270  
 DB 232 GEGEIIYAECSPEVIESFPQAVPTGFPFDVIS 286  
 RESULT 12  
 ABB76480  
 ID ABB76480 standard; Protein; 168 AA.  
 AC ABB76480;  
 XX  
 XX 23-SEP-2002 (first entry)  
 XX Novel human nitrilase-like protein.  
 XX Nitrilase; enzyme, human, NIP; gene therapy.  
 KW Homo sapiens.  
 XX

13-DEC-2001  
 25-JAN-2001; 2001WO 0826687  
 28 JAN-2000; 2000US 170000P  
 (LEXI-) LEXICON GENETICS INC.  
 Donoho G, Hilbun E, Scoville J, Turner CA, Friedrich G, Abuin A;  
 Zambrowicz B, Sands AT;  
 WPI: 2002-519061/55  
 N-PSDB; ABR83813.  
 Novel polynucleotides encoding novel human proteins sharing sequence  
 similarity with nitrilase proteins, useful for generating primers, and  
 probes which are used in identification of targets for drug discovery  
 Disclosure: Page 31; 36pp; English.  
 The present sequence is the protein sequence of a novel human  
 protein (NHP) that shares structural similarity with nitrilase  
 enzymes from a wide variety of organisms. The invention provides  
 nitrilase-like NHPs (see AB976477-83) and polynucleotides (see  
 ABR83808-15), host cell expression systems, fusion proteins,  
 antibodies, antagonists, agonists, and other compounds that  
 modulate NHP expression or activity. These can be used to  
 detect mutant or inappropriately expressed NHPs for the diagnosis  
 of disease, or to screen for drugs effective in the treatment of  
 cells and/or transgenic animals producing NHP. Engineered host  
 perturbations of the normal function of an NHP. Engineered host  
 cells and/or transgenic animals producing NHP can be used to  
 identify compounds that bind the endogenous NHP receptor and to  
 identify compounds that trigger NHP-mediated activities or  
 pathways.  
 Query Match 43.8%, Score 627, DP 23, Length 121, Mismatches 9, Indels 0, Gaps 0;  
 Best Local Similarity 100.0%; Pred. No. 9e-61;  
 Matches 121, Conservative 9, Mismatches 9, Indels 0, Gaps 0;  
 156 MFPAELAGTIVAGPSTGLVYVSPHLLTGGARWELLKSPAVNQQVAVATASAPDEAS 215  
 1 MPFAPLACTIVAGPSTGLVYVSPHLLTGGARWELLKSPAVNQQVAVATASAPDEAS 60  
 216 YVAMGHSVVNPNWGEVIAKAGSTHEAIVYSIDILKKLAEIRQQIPVFPKSDLYAVEMK 120  
 61 YVAMGHSVVNPNWGEVIAKAGSTHEAIVYSIDILKKLAEIRQQIPVFPKSDLYAVEMK 120  
 276 P 276  
 121 P 121  
 RESULT 14  
 ABR76481  
 ID ABR76481 standard; Protein; 130 AA  
 AC ABR76481;  
 XX ABR76481;  
 DT 21-SEP-2002 (first entry)  
 DE Novel human nitrilase-like protein.  
 KW Nitrilase; enzyme; human; NHP; gene therapy.  
 OS Homo sapiens.  
 XX WO200194566-A2.  
 XX 13-DEC-2001

25-JAN-2001; 2001WO 0826687  
 28 JAN-2000; 2000US 170000P  
 (LEXI-) LEXICON GENETICS INC.  
 Donoho G, Hilbun E, Scoville J, Turner CA, Friedrich G, Abuin A;  
 Zambrowicz B, Sands AT;  
 WPI: 2002-519061/55  
 N-PSDB; ABR83812.  
 Novel polynucleotides encoding novel human proteins sharing sequence  
 similarity with nitrilase proteins, useful for generating primers, and  
 probes which are used in identification of targets for drug discovery  
 Disclosure: Page 33; 36pp; English.  
 The present sequence is the protein sequence of a novel human  
 protein (NHP) that shares structural similarity with nitrilase  
 enzymes from a wide variety of organisms. The invention provides  
 nitrilase-like NHPs (see AB976477-83) and polynucleotides (see  
 ABR83808-15), host cell expression systems, fusion proteins,  
 antibodies, antagonists, agonists, and other compounds that  
 modulate NHP expression or activity. These can be used to  
 detect mutant or inappropriately expressed NHPs for the diagnosis  
 of disease, or to screen for drugs effective in the treatment of  
 cells and/or transgenic animals producing NHP. Engineered host  
 perturbations of the normal function of an NHP. Engineered host  
 cells and/or transgenic animals producing NHP can be used to  
 identify compounds that bind the endogenous NHP receptor and to  
 identify compounds that trigger NHP-mediated activities or  
 pathways.  
 Query Match 39.8%, Score 573, DP 23, Length 130;  
 Best Local Similarity 93.3%; Pred. No. 9e-55;  
 Mismatches 11, Conservative 9, Mismatches 9, Indels 0, Gaps 0;  
 156 MFPAELAGTIVAGPSTGLVYVSPHLLTGGARWELLKSPAVNQQVAVATASAPDEAS 215  
 1 MPFAPLACTIVAGPSTGLVYVSPHLLTGGARWELLKSPAVNQQVAVATASAPDEAS 60  
 216 YVAMGHSVVNPNWGEVIAKAGSTHEAIVYSIDILKKLAEIRQQIPVFPKSDLYAVEMK 274  
 61 YVAMGHSVVNPNWGEVIAKAGSTHEAIVYSIDILKKLAEIRQQIPVFPKSDLYAVEMK 274  
 RESULT 15  
 AAY68739  
 ID AAY68739 standard; Protein; 464 AA.  
 AC AAY68739;  
 XX AAY68739;  
 DT 05-MAY-2000 (first entry)  
 DE Amino acid sequence of the Nhl1 protein.  
 KW Nhl1 gene; nitrilase, tumor suppressor gene, PHT; chromosome 3p14.2;  
 FR3B; cancer; genome allele inactivation.  
 OS Homo sapiens.  
 OS Mus sp.  
 OS Drosophila melanogaster.  
 OS Caenorhabditis elegans.  
 XX Key Location/Qualifiers  
 PH Misc-difference 6  
 FT /note- "unspecified amino acid encoded by TNT"  
 XX WO200003685-A2.

Fri Jan 31 13:48:37 2003

PD 27-JAN-2000.  
XX  
XX  
PF 20-JUL-1999; 99WO-0516366.  
XX  
XX  
PR 20-JUL-1998; 98US-0093350.  
XX  
XX  
PA (OYJE-) UNIV JEFFERSON THOMAS.  
XX  
XX  
PI Croce CM;  
XX  
XX  
DR WPI; 2000-171195/15.  
DR N-PSDB; AAZ46101.  
XX  
XX  
PT Novel nitrilase homologs used as diagnostic and therapeutic reagents  
PT for the detection and treatment of cancer -  
XX  
XX  
PS Disclosure; Fig 6; 25pp; English.  
XX  
XX  
CC The present sequence is encoded by the coding region of human, murine,  
CC Drosophila melanogaster and Caenorhabditis elegans NIT1 gene. The  
CC human and mouse NIT1 genes are members of an uncharacterised  
CC mammalian gene family with homology to bacterial and plant nitrilases.  
CC The tumour suppressor gene FHIT in D. melanogaster and C. elegans code  
CC for fusion proteins in which the Fhit domain is fused with a Nit domain.  
CC In mouse and humans, FHIT and NIT are encoded by two different genes,  
CC localised on chromosomes 3 and 1 in human and 14 and 1 in mouse. The  
CC human FHIT gene at chromosome 3p14.2, spanning the constitutive  
CC chromosomal fragile site FRA3B, is often altered in most common forms  
CC of human cancer. The NIT1 protein overcomes the mutated inactivation  
CC of the genome alleles. The NIT1 genes, encoded polypeptides, derivatives  
CC and analogues of them, and antibodies are used as diagnostic and  
CC therapeutic reagents for the detection and treatment of cancers.  
XX  
XX  
SQ Sequence 464 AA;  
  
Query Match 31.8%; Score 485.5; DB 21; Length 464;  
Best Local Similarity 36.7%; Pred. No. 2.9e-44;  
Matches 105; Conservative 46; Mismatches 98; Indels 37; Gaps 6;  
  
QY 6 LALICLQISSKSDNVTACRFIREAATGGAKIVSLPCF-----NSPYGAK 52  
Db 84 VAVCVTSTIPURQNFETCAEIVPEAAPLGAFLPFAFTFIARDPAETLHLSEPLGK 143  
  
QY 53 YFPEYAEKIPGHSTOKLSEVAKECSIYLIIGGSIP-----EDAGKLYNTCAVFGPDGILL 107  
Db 144 LLEEV-----TQLARECGLWLSLGGFHERGQDWEQTKIYNCHVLLNSKGAVV 191  
  
QY 138 AYYPYTHLEFDIIVPGKITFQSEKTLSPGDSF-STFDTPTVCFVGLGICYDMRFAELAQIYA 166  
Db 132 ATYRATHLCDDVEIPGGNCESNTPGDSLESVPSTPAGKIGLAVCYDMRFPFELSALA 251  
  
QY 167 QPQGGTLLVYFPAFNITTPAHWELQRPAPVDNQVYVATASPA---RDDKASYVAWGHT 223  
Db 252 QAGAILLAYSAPGSITTPAHWEVLI PAPAIFETCYVVAQAACGPHHFKRASY---GHSM 308  
  
QY 224 VVNFQGEVLAKAGTEEAIVYSIDLEKLAIEIQQIPVPQKPSDLY 269  
Db 309 VVDFWGTVMARCSGDPGLTAPIDLNLYIQKPPHLPVFQHPRPDLY 354

Search completed: January 31, 2003, 07:51:17  
Job time : 39 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	1358	99.9	1259	24	AAD25456
2	924	97.9	1001	24	AAAF4414
3	831	83.1	831	24	ABN3808
4	796	59.6	859	24	ABN3814
5	702.2	51.7	1725	21	BAC77116
6	679.8	50.0	1092	24	BAC35457
7	508.2	37.4	3093	24	ABN3815
8	472.2	34.7	489	24	ABN04371
9	470.2	34.0	453	24	ABN2557















[illegible]











## RESULT 14

ABN83812  
ID ABN83812 standard, cDNA, 39: BP.

XX ABN83812;

XX 23-SEP-2002 (first entry)

XX Polynucleotide encoding novel human nitrilase-like protein.

XX Nitrilase, enzyme, human, NHP, gene therapy; gene, ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH 1..393  
FT CDS /\*tag= a  
/\*product= "NHP"

PN WO200194566-A2.

XX 13-DEC-2001.

XX 25 JAN 2001; 2001WO-US26687.

XX 29 JAN 2000; 2000US-170000P.

XX (LEXI-) LEXICON GENETICS INC.

XX Donoho G, Hilbun E, Scoville J, Turner CA, Friedrich G, Abuin A;  
PI Zambrowicz B, Sands AT;

XX WPI; 2002-519061/55.

XX P-PSDB; ABN76481.

XX Novel polynucleotides encoding novel human proteins sharing sequence  
PT similarity with nitrilase proteins, useful for generating primers, and  
PT probes which are used in identification of targets for drug discovery

XX Disclosure; Page 32; 36pp; English.

XX The present sequence is the coding sequence of an isolated nucleic  
XX acid molecule encoding a novel human protein (NHP) that shares  
XX structural similarity with nitrilase enzymes from a wide variety of  
XX organisms. The invention provides nitrilase-like NHPs (see  
XX ABN76477-83) and polynucleotides encoding them (see ABN83808-15),  
XX host cell expression systems, fusion proteins, antibodies,  
XX antagonists, agonists, and other compounds that modulate NHP  
XX expression or activity and are useful for diagnosis, drug screening,  
XX clinical trial monitoring and the treatment of physiological  
XX disorders. The NHP polynucleotide sequences were compiled from  
XX gene trapped cDNAs and clones isolated from human prostate, lymph  
XX node, pituitary, mammary gland and kidney cDNA libraries. NHP  
XX polynucleotide sequences can be used in the treatment of diseases  
XX involving NHP, for detecting mutant NHPs or inappropriately  
XX expressed NHPs for the diagnosis of disease, and for screening drugs  
XX effective in treatment of symptomatic or phenotypic manifestations  
XX perturbing the normal function of NHP in the body. Nucleic acid  
XX fragments can be used as primers in PCR to screen libraries, isolate  
XX clones and prepare cloning and sequencing templates. They can also  
XX be used as hybridisation probes for screening libraries and for  
XX assessing gene expression patterns, as well as for identification,  
XX selection and validation of novel molecular targets for drug  
XX discovery. Labeled NHP probes can be used to screen a genomic  
XX library to identify polymorphisms, determine the genomic structure of  
XX a given locus/allele and design diagnostic tests. The probes also  
XX have use in defining and monitoring drug action and toxicity.  
XX NHP oligonucleotides can be used to design antisense, ribozyme  
XX and/or triple helix molecules. Addressable arrays  
XX comprising the NHP polynucleotides can be used to identify and

CC characterise the tissue expression of a gene and to provide detailed  
CC information about transcriptional changes involved in a specific  
CC pathway, potentially leading to the identification of novel  
CC components or gene functions that manifest themselves as novel  
CC phenotypes.

XX Sequence 393 BP; 107 A; 93 C; 104 G; 89 T; 0 other;

Query Match 24.4%, Score 321.4; DB 24, Length 393;

Best Local Similarity 95.5%; Pred. No. 9.6e-75;

Matches 341; Conservative 0; Mismatches 16; Indels 0; Gaps 0.

QY 497 ATGCGCGTTTGCAGAGCTTGCACAAATTTACGACACAGAGAGGCTGCCAGGCTGTGTATAT 546  
|||||  
Db 1 ATGCGCGTTTGCAGAGCTTGCACAAATTTACGACACAGAGAGGCTGCCAGGCTGTGTATAT 60

QY 547 CCAGAGAGCTTTTAAATTTAACTATGAGACAGAGAGGCTATGAGAGAGGCTG 606  
|||||  
Db 61 CCAGAGAGCTTTTAAATTTAACTATGAGACAGAGAGGCTATGAGAGAGGCTG 120

QY 607 GGTGTTGATATATCAGGTGTATGTGACACAGAGAGGCTATGAGAGAGGCTG 666  
|||||  
Db 121 GGTGTTGATATATCAGGTGTATGTGACACAGAGAGGCTATGAGAGAGGCTG 180

QY 667 TATGTTGCTTGGGGACACAGAGAGGCTGTGACAGAGAGGCTGTGAGAGGCT 726  
|||||  
Db 181 TATGTTGCTTGGGGACACAGAGAGGCTGTGACAGAGAGGCTGTGAGAGGCT 240

QY 727 GGCACAGAGAGAGAGAGAGGCTGTGATTCAGACATAGAGCTGAAGAGGCTG 786  
|||||  
Db 241 GGCACAGAGAGAGAGAGAGGCTGTGATTCAGACATAGAGCTGAAGAGGCTG 300

QY 787 CAGCAAAATCCCGGTTTTTAGACAGAGAGGCTATGAGAGGCTG 843  
|||||  
Db 301 CAGCAAAATCCCGGTTTTTAGACAGAGAGGCTATGAGAGGCTG 900

## RESULT 15

AAC02370

ID AAC02370 standard; cDNA; 361 BP.

XX AAC02370;

XX 06-OCT-2000 (first entry)

XX Human secreted protein 5' EST, SEQ ID NO: 2368.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB 2000; 2000EP-0200610.

XX 26 FEB 1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX P-PSDB; AAG02364.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX Claim 1; SEQ ID 2368; 71pp + CD-ROM; English.

